Slowness and Sparseness for Unsupervised Learning of Spatial and Object Codes from Naturalistic Data

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Biologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
Humboldt-Universität zu Berlin

von
Herr Dipl.-Inf. Mathias Franzius
geboren am 6.3.1975 in Minden

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Christoph Markschies

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Dr. Christian Limberg

Gutachter:

1. Prof. Dr. L. Wiskott
2. Dr. R. Kempter
3. Prof. Dr. A. Treves

eingereicht am:	3. 12. 2007
Tag der mündlichen Prüfung:	30. 5. 2008

Abstract

This thesis introduces a hierarchical model for unsupervised learning from naturalistic video sequences. The model is based on the principles of slowness and sparseness. Different approaches and implementations for these principles are discussed. A variety of neuron classes in the hippocampal formation of rodents and primates codes for different aspects of space surrounding the animal, including place cells, head direction cells, spatial view cells and grid cells. In the main part of this thesis, video sequences from a virtual reality environment are used for training the hierarchical model. The behavior of most known hippocampal neuron types coding for space are reproduced by this model. The type of representations generated by the model is mostly determined by the movement statistics of the simulated animal. The model approach is not limited to spatial coding. An application of the model to invariant object recognition is described, where artificial clusters of spheres or rendered fish are presented to the model. The resulting representations allow a simple extraction of the identity of the object presented as well as of its position and viewing angle.

Keywords:
Hippocampus, Object Recognition, Place Cells, Unsupervised Learning

Zusammenfassung

Diese Doktorarbeit führt ein hierarchisches Modell für das unüberwachte Lernen aus quasi-natürlichen Videosequenzen ein. Das Modell basiert auf den Lernprinzipien der Langsamkeit und Spärlichkeit, für die verschiedene Ansätze und Implementierungen vorgestellt werden. Eine Vielzahl von Neuronentypen im Hippocampus von Nagern und Primaten kodiert verschiedene Aspekte der räumlichen Umgebung eines Tieres. Dazu gehören Ortszellen (place cells), Kopfrichtungszellen (head direction cells), Raumansichtszellen (spatial view cells) und Gitterzellen (grid cells). Die Hauptergebnisse dieser Arbeit basieren auf dem Training des hierarchischen Modells mit Videosequenzen aus einer Virtual-Reality-Umgebung. Das Modell reproduziert die wichtigsten räumlichen Codes aus dem Hippocampus. Die Art der erzeugten Repräsentationen hängt hauptsächlich von der Bewegungsstatistik des simulierten Tieres ab. Das vorgestellte Modell wird außerdem auf das Problem der invaranten Objekterkennung angewandt, indem Videosequenzen von simulierten Kugelhaufen oder Fischen als Stimuli genutzt wurden. Die resultierenden Modellrepräsentationen erlauben das unabhängige Auslesen von Objektidentität, Position und Rotationswinkel im Raum.

Schlagwörter:
Hippocampus, Objekterkennung, Ortszellen, Unüberwachtes Lernen

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The brain has no direct access to the world around it. In order to find out about its environment, it has to rely on electrical impulses generated by eyes, ears, and other sensory organs. The most important information an animal needs to retrieve from these sensory inputs includes its own position ("Where am I?") and heading direction ("In what direction am I looking?") as well as the position, identity and viewing angle of objects surrounding the animal ("What object do I see? Where is it?"). This information forms the basis for the ability to navigate in the environment and manipulate objects therein.

This thesis introduces a model that allows the extraction of all these fundamental types of information from realistic quasi-natural video sequences under certain constraints. It is based on three articles, two of which are already peer-reviewed and published. The first article by Mathias Franzius, Roland Vollgraf, and Laurenz Wiskott titled *From Grids to Places* was published in the Journal of Computational Neuroscience [Franzius et al., 2007b] and is integrated into Chapter 2. The second article by Mathias Franzius, Henning Sprekeler, and Laurenz Wiskott titled *Slowness and Sparseness Lead to Place, Head-Direction and Spatial-View Cells* was published by PLoS Computational Biology [Franzius et al., 2007a] and forms the basis for Chapter 4. A third article by Mathias Franzius, Niko Wilbert, and Laurenz Wiskott is in preparation [Franzius et al., 2007c] and outlined in Chapter 5.

A biologically realistic model for spatial coding and invariant object recognition in the brain can advance the understanding of two major scientific fields. Such a model can help identify plausible functional principles underlying the highly complex "neural implementation" of these processes in the brain and make testable predictions for experiments. Furthermore, advances in object recognition and spatial coding can help to improve computer vision and possibly applications in robotics. However, the applicability of the computational model presented here in "real life" robotics needs to be shown in the future.

1.1 Overview

This thesis is structured in six chapters. In Chapter 2, the slowness and sparseness principles are introduced, together with a motivation, history, and models of these unsupervised learning rules. These two principles form the basis of the model presented in later chapters. Chapter 3 commences with a short introduction to approaches for navigation and continues with an overview of the major types of neurons coding for spatial features in the hippocampal formation. Chapter 4 contains the major results of this thesis. A hierarchical model for unsupervised learning of spatial codes from quasi-natural videos is introduced. This model reproduces the characteristics of most spatially coding neuron types in the brain known so far. In Chapter 5, a similar model is applied to the domain of invariant object recognition for complex three-dimensional objects under translation and in-depth rotation. This model creates view-invariant representations of objects similar to those in the inferotemporal cortex. At the same time, information about position and viewing angle is extracted. Finally, Chapter 6 summarizes the thesis, discusses the main advantages and shortcomings and gives an outlook on possible future work.

Chapter 2

The Principles of Temporal Slowness and Sparseness

Most results of this thesis are based on principles of unsupervised learning. Such methods find representations of aspects of their inputs without an external teaching or supervision signal that explicitly defines a desired result. Often some internally generated performance measure is available instead of an external teaching signal that is used to guide the training process. Such a performance measure (cost function, objective function) can be used to derive a learning rule, which is applied to adapt connection weights during a training phase [Becker and Zemel, 2003].
This chapter introduces two families of unsupervised learning rules. A third approach, Principal Component Analysis, is introduced in the first section since it constitutes an important component of most of the following learning rules. In Section 2.2, learning rules based on the *slowness principle* are introduced. Section 2.3 explains the concept and different implementations of *sparse coding*. Although slowness and sparseness goal functions seem to be highly different approaches, for some special cases a close relationship exists. Blaschke et al. [2006] have proven analytically the identity of linear SFA with ICA under certain conditions.
The main results in this thesis (Chapters 4 and 5) are based on the slowness principle, namely nonlinear SFA. The results in Chapter 4 are additionally based on a final linear sparse coding step.

2.1 Principal Component Analysis

One of the most popular techniques of unsupervised learning is Principal Component Analysis (PCA), which finds a rotated coordinate system such that the input data representation in the new coordinate system is decorrelated (cf. Figure 2.1). Additionally, the basis vectors are sorted by decreasing variance of the input data in these directions. A typical application of PCA

is dimensionality reduction. Consider, for example, a two-dimensional data cloud embedded in an n-dimensional space with $n \gg 2$ and subjected to a small amount of additive noise. In order to represent the data in the high-dimensional space one n-dimensional vector per data point is necessary. But if the extension of the data points in all other directions is minimal, for example because of small noise variance, only the subspace of high variance is of interest. As this subspace is only two-dimensional, each data point can be represented by a vector $v \in \mathbb{R}^2$, using only $\frac{2}{n}$ of the storage space of the original representation[1]. This compression is optimal in the sense of least-square reconstruction errors in a linear model [Jolliffe, 2002].
Principal Component Analysis can be computed iteratively with an *online learning rule*, for example by extensions of Oja's Rule [Oja, 1982, 1992] in an artificial neural network. Here, each data point causes a little change of the weight vectors such that on average their angles to the true principal directions are reduced. Alternatively, PCA can be computed by diagonalizing the covariance matrix of the input data. The eigenvectors with largest eigenvalues point in the direction of highest variance in the data distribution and input data projected on these vectors is decorrelated. This approach belongs to the class of *offline* or *batch learning rules* where all data has to be presented before any output is generated. In mathematical terms, this approach finds a linear transformation $\mathbf{A}$ from an input vector $\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T$ to an output vector $\mathbf{y}(t) = (y_1(t), y_2(t), \ldots, y_k(t))^T = \mathbf{A}\mathbf{x}(t)$. $\mathbf{A}$ can be found by diagonalizing the covariance matrix $\mathbf{C} = \langle (\mathbf{x}(t) - \langle \mathbf{x}(t) \rangle_t)(\mathbf{x}(t) - \langle \mathbf{x}(t) \rangle_t)^T \rangle_t$: $\mathbf{C}v_i = \lambda_i v_i$, where $\langle \cdot \rangle_t$ denotes temporal averaging, v_i are the eigenvectors, and λ_i the corresponding eigenvalues of $\mathbf{C}$. If the eigenvalues are distinct, a unique orthogonal basis can be built from the eigenvectors by ordering them by decreasing eigenvalues. In this basis, the first axis (v_1) points into the direction of largest variance of the input data, and the last axis (v_n) points into the direction of smallest variance. If the eigenvalues are not distinct, any arbitrary rotation in the subspace(s) spanned by the eigenvalues with corresponding identical eigenvalues is a valid solution of PCA. If dimensionality reduction is performed with PCA, as in the example above, only the first k high-variance dimensions are kept, reducing the transformation matrix $\mathbf{A}$ to dimensionality $n \times k$.

Often, offline learning rules are more efficient in terms of computation time and more demanding in terms of computer memory than online learning rules. In the case of PCA, both approaches eventually yield the same result[2], but convergence of the online approach can be extremely slow. On the other hand, online rules are often considered more plausible as models of neural function than offline rules. Nevertheless, both approaches are only different implementations of the same principle. In this thesis, a less biologically

[1] Additionally, the $n \times 2$ transformation matrix needs to be stored.
[2] up to arbitrary rotations in subspaces of (nearly) identical variances

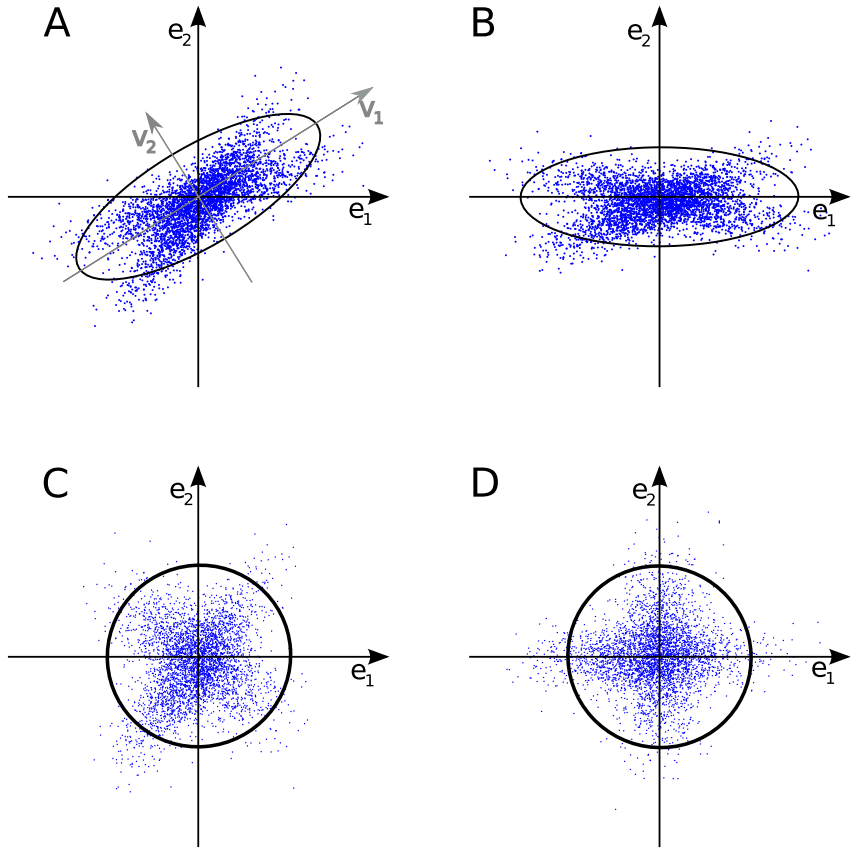

Figure 2.1: Illustration of Principal Component Analysis (PCA) and Independent Component Analysis (ICA). A two-dimensional input data distribution is plotted in panel A. The gray ellipse visualizes the covariance of the data. Principal Component Analysis rotates the principal axes of the ellipse (v_1 and v_2) onto the Euclidean axes e_1 and e_2 (panel B). Whitening shrinks or expands the data along the principal axes such that the variance along all axes is one (panel C). The covariance matrix of the data distribution in panel C is now the identity matrix. Thus, any rotation of the data yields the same covariance structure. As PCA ignores higher order statistics, such rotations remain free parameters. In panel D, the data has been rotated by ICA such that the projections on the Euclidean axes are statistically maximally independent. The choice of a measure of statistical independence differs between ICA algorithms.

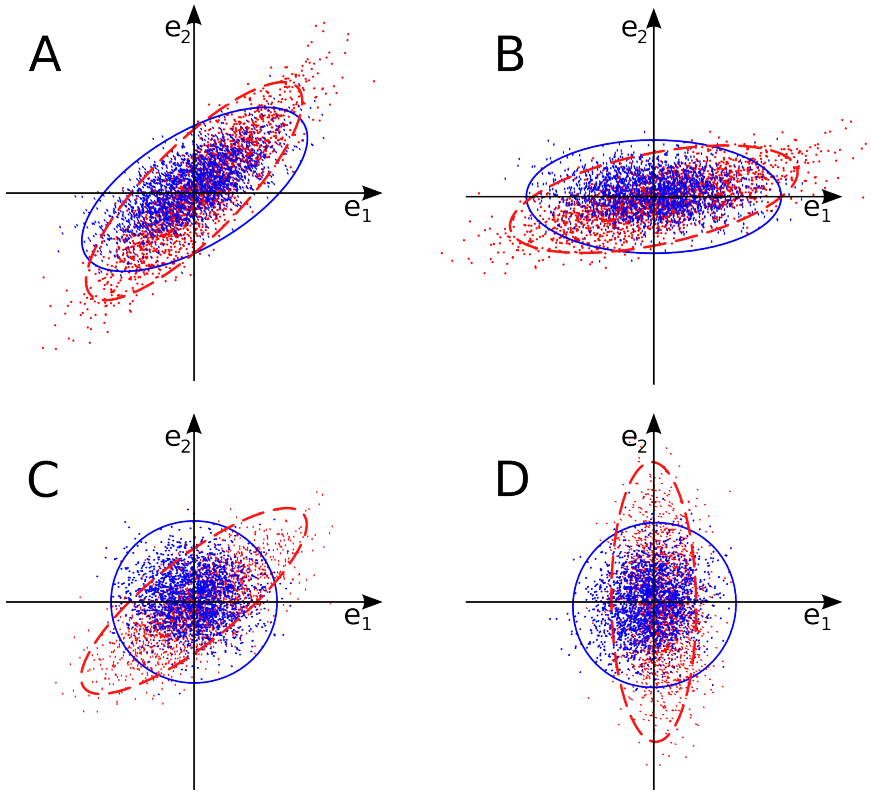

Figure 2.2: Illustration of Slow Feature Analysis (SFA). A two-dimensional input data distribution is plotted in panel A as a cloud of blue points and the time derivatives of the data are plotted as red points. The ellipses illustrate the covariance of data and time derivatives. Panel B shows the effect of Principal Component Analysis on the data cloud. The principal axes of the data ellipse are aligned with the Euclidean axes and the derivatives are accordingly rotated by the same amount. In panel C, the data has been whitened, causing a similar deformation to the derivatives. In contrast to ICA (see Figure 2.1), the remaining degrees of freedom for rotations in whitened space are chosen such that the principal axes of the derivative's covariance ellipsoid (red) are aligned with the Euclidean axes in decreasing order. Hence, the projection of the input data onto the first axis $e1$ has unit variance and is as slow as possible. The projections onto the other axes also have unit variance, are mutually decorrelated, and are ordered by decreasing slowness.

plausible implementation of an otherwise plausible principle will often be preferred if it is computationally more efficient or if it allows a more direct and possibly analytical description of the results.

2.2 Slowness

PCA is insensitive to the temporal structure of the input data, i.e., a temporally permuted presentation of the input yields the same results. But often the temporal structure of data contains important information: in many scenarios, stimuli that typically occur temporally close should elicit similar outputs. This similarity measure constitutes the basis for the objective functions necessary for unsupervised learning rules as mentioned above. The *slowness principle* gives rise to learning rules that are sensitive to the temporal structure of input data. In Section 2.2, different versions of unsupervised learning rules for optimizing the slowness principle are presented. Most of these rules are online approaches, whereas Slow Feature Analysis (SFA) is an offline approach.

2.2.1 Motivation

When we observe a sequence of different views of an object – for example from different sides, we usually perceive that the identity of the object does not change. The visual input on the other hand can change dramatically even when the object is only slightly turned or moved. For the extreme example of a zebra, a little movement of the animal or of the observer will change the light intensity of most individual retinal receptor from white to black or vice versa, although the observer might not perceive any relevant change of the environment. Generally, sensory data changes on a shorter timescale than the behaviorally relevant configuration of the environment (see Figure 2.3). However, this observation only holds *on average* over long timescales on the order of minutes or hours. Sometimes relevant changes can occur very rapidly and it is highly important for an efficient sensory system to capture a changed configuration as quickly as possible. The appearance of a predator in the field of view or of a large truck heading towards you, for example, are behaviorally highly relevant cases where a slow sensory system can act as a strong negative evolutionary selection factor. Thus we are confronted with an apparent paradoxon: on the one hand we would like to identify features that typically vary slowly in the hope to obtain a useful representation of the environment. On the other hand we need to identify these features from a given sensory signal as quickly as possible. In mathematical terms, the quickest possible system reacts instantaneously, that is, it yields an output without any delay at the presentation of a 'stimulus snapshot', e.g., a single static picture. In the context of spiking neurons of the nervous system, the typical latencies are on the order of 10 ms. The processing

latencies in the primate visual system during object recognition tasks, for example, are in the range of 10–30 ms per cortical area, suggesting that no extensive recurrent processing takes place for this task [Thorpe et al., 1996, Rolls, 2007]. The constraint of instantaneous – or at least quick – response rules out the trivial approach of low-pass filtering to find slowly varying features. A low-pass filtered signal changes slower than the unfiltered signal but usually discards relevant information.

A system that extracts aspects of its inputs that vary slowly or rarely has to disregard aspects that vary on quicker timescales. If the stimulus variance is caused by a limited number of transformations (e.g., rotation and translation of an object), the system will code for those transformations that change most slowly or most seldom while it becomes *invariant* to the other transformations[3]. Learning of invariant representations is an important topic for many research areas involving sensory coding, including spatial coding, object recognition, face and speech recognition.

Slowly changing configurations of the environment might include the position, angle, and identity of objects surrounding us. An application of the slowness principle on these topic are described in Chapter 5. Alternatively, slowly changing configurations might be given by the observer's own position and viewing direction in space, which is the topic of Chapter 4.
Although slowness is a necessary feature of many high-level cognitive codes, slowness alone is in general not sufficient to find the "most interesting" representation from sensory data. In this sense, the slowness principle is a *heuristic* that reduces the search space of possible meaningful representations. Additional constraints might be necessary in order to find a specific representation. Such constraints can be incorporated, among others, by priors in probabilistic models (cf. Section 2.2.2) or an adaptation of the learning rate (see Section 4.1.2).

The slowness principle is connected to the principle of predictive coding, which postulates that sensory systems are adapted to the statistics of their inputs in a way that maximizes information about future inputs. As a representation learned by the slowness principle typically changes only little over short time scales, its value at any given time point often is similar to its value in the close future and thus informative about the state of the environment in the future. This idea is elaborated in Creutzig and Sprekeler [2008] where the equivalence of the two principles is shown under the constraints of a linear system and a reversible Gaussian signal statistics.

2.2.2 Approaches for Slowness Learning

The *slowness principle* forms the basis for a variety of learning rules [e.g., Földiák, 1991, Mitchison, 1991, Stone and Bray, 1995, Wiskott and Se-

[3]This only holds if the computational power of the system is sufficient.

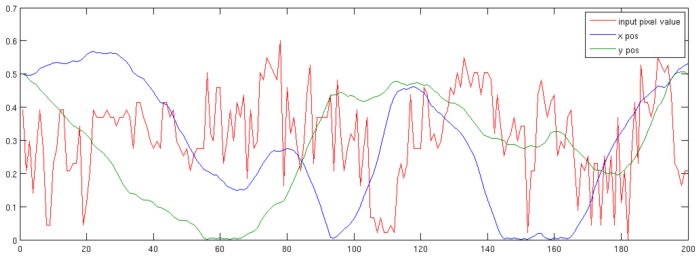

Figure 2.3: Illustration of raw sensory data and extracted slow features. The value of a single fixed gray scale pixel from an input movie sequence of a virtual rat moving in a virtual room (cf. Chapter 4) is plotted in red over 200 time steps. Many thousands of these pixel traces constitute the input video and all necessary data is contained in these data to estimate the position of the rat within the simulated room. However, the extraction of these two coordinates (x- and y-position of the rat in the room), as plotted in blue and green, is not trivial. One evident difference between the red and the blue trace is the temporal variability. The red signal changes much more quickly than the blue one.

jnowski, 2002, Hurri and Hyvärinen, 2003, Körding et al., 2004, Berkes and Wiskott, 2005]. Some authors refer to this principle as *temporal stability* or *temporal coherence*, depending on the actual implementation. A short review of the most important variants is given below, a more detailed discussion of different slowness functions can be found in [Berkes, 2005b].

Slow Feature Analysis

The implementation of the slowness principle used in this thesis is *Slow Feature Analysis* (SFA) as introduced by Wiskott [Wiskott, 1998, Wiskott and Sejnowski, 2002]. Slow Feature Analysis solves the following learning task: Given a multidimensional input signal we want to find instantaneous scalar input-output functions that generate output signals that vary as slowly as possible but still carry significant information. To ensure the latter we require the output signals to be uncorrelated and have unit variance. In mathematical terms, this can be stated as follows:
Optimization problem: *Given a function space $\mathcal{F}$ and an I-dimensional input signal $\boldsymbol{x}(t)$ find a set of J real-valued input-output functions $g_j(\boldsymbol{x}) \in \mathcal{F}$ such that the output signals $y_j(t) := g_j(\boldsymbol{x}(t))$*

$$\textit{minimize} \quad \Delta(y_j) := \langle \dot{y}_j^2 \rangle_t \tag{2.1}$$

under the constraints

$$\langle y_j \rangle_t = 0 \quad \textit{(zero mean)}, \tag{2.2}$$

$$\langle y_j^2 \rangle_t = 1 \quad \textit{(unit variance)}, \tag{2.3}$$

$$\forall i < j : \langle y_i y_j \rangle_t = 0 \quad \textit{(decorrelation and order)}, \tag{2.4}$$

with $\langle \cdot \rangle_t$ and $\dot{y}$ indicating temporal averaging and the derivative of y, respectively.

Equation (2.1) introduces the Δ-value, which is a measure of the temporal slowness of the signal $y(t)$. It is given by the mean square of the signal's temporal derivative, so small Δ-values indicate slowly varying signals. The constraints (2.2) and (2.3) avoid the trivial constant solution and constraint (2.4) ensures that different functions g_j code for different aspects of the input.

It is important to note that although the objective is slowness, the functions g_j are instantaneous functions of the input, so that slowness cannot be enforced by low-pass filtering. Slow output signals can only be obtained if the input signal contains slowly varying features that can be extracted instantaneously by the functions g_j.

In the computationally relevant case where $\mathcal{F}$ is finite-dimensional the solution to the optimization problem can be found by means of Slow Feature Analysis [Wiskott and Sejnowski, 2002, Berkes and Wiskott, 2005]. This algorithm, which is based on an eigenvector approach, is guaranteed to find the global optimum. Biologically more plausible learning rules for the optimization problem, both for graded response and spiking units exist [Hashimoto, 2003, Sprekeler et al., 2007].

If the function space is infinite-dimensional, the problem requires variational calculus and will in general be difficult to solve. In Section 4.2 we demonstrate that the optimization problem for high-dimensional visual input used in Chapter 4 can be reformulated for a low-dimensional configural input of position and orientation. In this case, the variational calculus approach becomes tractable and allows to make analytical predictions for the behavior of the full model. The full analytical treatment is given in [Franzius et al., 2007a].

SFA is equivalent to maximum likelihood learning in a linear Gaussian state-space model with an independent Markovian prior [Turner and Sahani, 2007]. In this framework, SFA is a deterministic special case of a probabilistic model and can be extended by various well-known techniques from this field. The advantages of the approach by Turner and Sahani [2007] include its formulation as a generative model, the ability to cope with missing data, and with high measurement noise. This approach, however, assumes a linear and invertible mapping and for this special case SFA can be used as a generative model as well. Furthermore the probabilistic formulation by

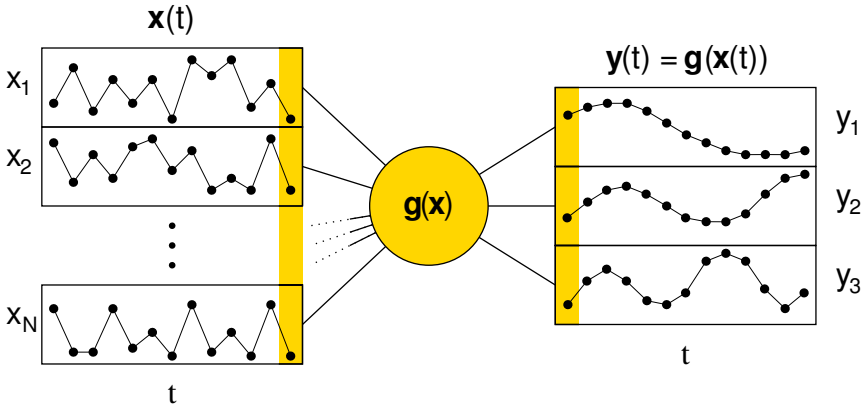

Figure 2.4: Illustration of the optimization performed by SFA. A quickly varying multidimensional input signal $\mathbf{x}(t)$ is instantaneously transformed into a slowly varying multidimensional output signal $\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t))$ by a set of transfer functions $\mathbf{g}(t)$. Figure courtesy of Laurenz Wiskott.

Turner and Sahani is computationally much more demanding and possibly restricted to low-dimensional problems.

Other Approaches for Slowness Learning

Many approaches for invariant coding are either supervised, or use explicitly built-in invariances [e.g., the Neocognitron by Fukushima, 1980], whereas the approaches in this section are unsupervised and only based on the statistics of the input data[4].

The model by Földiák adapts the weights according to the *trace rule*, which is a modified Hebbian learning rule [Földiák, 1991]. The trace rule updates a neuron's input weight vector $\mathbf{w}$ proportionally to the product of the neuron's current input $\mathbf{x}$ and the neuron's *trace value* $\bar{y}^\tau$ according to the rule: $\delta w_j^\tau = \alpha \bar{y}^\tau x_j^\tau$, where α is the learning rate with $0 \leq \alpha \leq 1$. The trace value $\bar{y}^\tau$ at time step τ is defined as the exponentially weighted mean activity of the neuron in the past:

$$\bar{y}^\tau := (1-\eta)y^\tau + \eta \bar{y}^{\tau-1}, \tag{2.5}$$

where $\bar{y}^{\tau-1}$ is the trace value of the prior time step and η (with $0 \leq \eta < 1$) defines how much past activities influence the trace. The weight vector $\mathbf{w}$ has to be normalized explicitly or implicitly in order to prevent infinite growth. Földiák's model was inspired by the visual systems's ability to achieve invariant recognition despite changes of viewing angle, eye position,

[4] A possible exception is the model by Mitchison below, which optionally includes a supervised bias.

distance, size, orientation etc. and designed to improve the Neocognitron [Fukushima, 1980]. Shift invariance in the original Neocognitron model was explicitly built-in by a weight-sharing system and a fixed pooling step. The weight-sharing guarantees that in a given layer of a hierarchical network the same feature is extracted at all possible locations and thus the same feature is passed on to the higher layer independently of the stimulus location. Földiák's model improves this approach by learning translation invariance with the trace rule instead of forcing the invariance explicitly. The type of invariance learned by the system, however, now depends on the transformation statistics of the stimuli.
The trace rule was later applied in many other models for face and object recognition [e.g., Rolls, 1992, Wallis and Rolls, 1997, Rolls and Stringer, 2006, Rolls and Deco, 2002]. The VisNet model as the most prominent model that uses the trace rule is discussed in more detail in Chapter 5.3.1. The relation between the trace rule and SFA is discussed in detail in [Sprekeler et al., 2007].

Mitchison [1991] derived another learning rule based on the slowness principle by means of gradient descent on the goal function $\Psi = \langle \dot{y}^2 \rangle$, which is identical to the formulation by SFA in Equation 2.1, except for variance normalization. For a linear unit $w = \sum w_i x_i$ the gradient descent then has the form of an anti-Hebbian rule: $\Delta w_i = -\alpha \Delta x_i \Delta y$, where Δ is the discrete approximation of the temporal derivative. Explicit weight normalization is applied to prevent the weight vector from converging to zero. As this approach so far only describes the learning of a single output unit, an additional bias mechanism is introduced in order to influence different units towards different outputs. This bias is used to integrate prior knowledge into the learned representations by a supervision signal but could also be used to learn a set of decorrelated representations in an unsupervised manner.

The model by Stone and Bray [1995] approaches the problem of finding slowly varying features while avoiding the trivial solution of constant signals (e.g., by weights decaying to zero) without explicit weight normalization by introducing two different time scales. Maximizing the objective function Ψ minimizes the short-term variance U of an output y but also maximizes their long-term variance V:

$$\Psi := \frac{1}{2} \log \frac{V}{U} = \frac{1}{2} \log \left(\frac{\sum_t (y(t) - \bar{y}(t))^2}{\sum_t (y(t) - \tilde{y}(t))^2} \right),$$

where both $\bar{y}$ and $\tilde{y}$ are exponentially weighted temporal sums of the output y. The timescale of $\bar{y}$ is longer than that of $\tilde{y}$ (i.e., $\eta_U < \eta_V$ as defined for the trace rule above) but otherwise both resemble traces as defined by the trace rule in Equation 2.5. In the linear case the weights are optimized by

gradient descent as follows:

$$\Delta w_j = \alpha \frac{\partial \Psi}{\partial w_j} = \frac{\alpha}{V} \langle (y - \bar{y})(x_j - \bar{x}_j) \rangle - \frac{1}{U} \langle (y - \tilde{y})(x_j - \tilde{x}_j) \rangle \ . \tag{2.6}$$

Multiple outputs can be learned by using an additional asymmetric decorrelation term. Unlike SFA, this goal function is insensitive to the output signal variance and during optimization the variance will likely float. Otherwise, for very long timescales of $\bar{y}$ and very short timescales of $\tilde{y}$, this approach becomes similar to SFA (see below).
This model was used to learn a transformation of a binary local code (i.e., a single active element in a one- or two-dimensional input) into a graded distributed code representing the active element's coordinate(s). In Bray and Martinez [2002], this approach is reformulated as a kernel-based nonlinear version with the intention of circumventing the curse of dimensionality that SFA with explicit expansion suffers from. The application of the kernel method instead of explicit expansion, however, shifts the computational complexity from size of the feature space to the number of support vectors. For large training sets, an adequate choice of only few support vectors is crucial to avoid the curse of dimensionality [e.g., Schoelkopf et al., 1999][5].
Alternatively to the gradient descent optimization in the original publication, the problem can also be solved as a generalized eigenvalue problem like SFA. In contrast to SFA, the covariance matrix $\mathbf{C}_U$ of the short-term trace $\tilde{y}$ (instead of the covariances of the derivatives) and the covariance matrix $\mathbf{C}_V$ of the long-term trace $\bar{y}$ (instead of the total covariances) are used. The solutions $\mathbf{W}$ are found, as in SFA, by solving the equation $\mathbf{C}_U\mathbf{W} = \mathbf{C}_V\mathbf{W}\Lambda$ where $\mathbf{W}$ is the transformation matrix consisting of generalized eigenvectors and Λ is the diagonal matrix of generalized eigenvalues. Thus this approach is a generalization of SFA.

König and colleagues introduced yet a different formulation of the slowness principle and applied it to natural images in order to model complex cells in primary visual cortex [Körding et al., 2004, Kayser et al., 2001], for modeling object recognition [Franzius, 2003, Einhäuser et al., 2005], and hippocampal place cells [Wyss et al., 2006]. This formulation is based on gradient descent on the goal function

$$\Psi = \sum_i \frac{\langle \dot{y}_i^2 \rangle_t}{\mathrm{var}_i(y_i)} + \alpha \sum_{i \neq j} \frac{\langle y_i y_j \rangle_t^2}{\langle y_i^2 \rangle_t \langle y_j^2 \rangle_t} + \beta \sum_i \langle y_i \rangle_t \tag{2.7}$$

and variations thereof. The first term enforces slowly varying outputs, similarly to the SFA formulation or to that by Mitchison [1991]. Similar to

[5] As in this thesis typically many more data points than data dimensions are used, hierarchical SFA with explicit polynomial expansion is applied instead of kernel expansion. The results in Chapters 4 and 5 as well as in [Berkes, 2005b] should disprove the claim by Bray and Martinez [2002] that SFA is "limited to simple theoretical simulations".

the formulation by Stone and Bray [1995] in Equation 2.6, this term is invariant with respect to the output variance. A notable difference to other approaches is, however, that the *average* slowness (i.e., the sum of the delta values) of the outputs y_i is minimized and thus the result is unique only up to orthogonal transformations. This also means that there is no order of the solutions – instead all y_i typically have roughly the same slowness (unpublished observation). The second term of the goal function enforces a soft decorrelation of the y_i. The relative strength of decorrelation vs. slowness is governed by the trade-off parameter α. The third term enforces mean activities close to zero. In alternative formulations [e.g., Körding et al., 2004] this term is dropped and instead in the first two terms the mean output is explicitly subtracted. The models of this group employ nonlinear transformations that are generally a sum of linear filters whose output has been subject to a point nonlinearity, which typically is squaring.
The application of this learning rule in a model for place cell learning will be discussed in in Chapter 4 and the application in a model for object recognition in Chapter 5

2.3 Sparseness

As we have seen before, PCA finds a data representation where individual dimensions are uncorrelated, i.e., the cross-correlations of the data distribution vanish. But in most cases decorrelation of data does not result in statistically independent representations because higher order (cross-) cumulants are ignored by PCA. Approaches to reduce these dependencies in order to find independent representations are called Independent Component Analysis (ICA). Many ICA algorithms search for results with extremal higher statistical cumulants or moments (e.g., kurtosis). Such representations with maximal higher cumulants are typically often inactive and seldom highly active[6]. This property is called *sparseness*. ICA is only one of the approaches for finding sparse codes discussed in this section. Alternative approaches, including Competitive Learning (CL), are introduced as well.
Sparse neuronal codes occur in many cortical areas. Among these are many spatial codes, mainly from neurons in the hippocampal formation, which are discussed in Chapter 3. Other neurons with sparse codes can be found in higher visual and multimodal cortices like the inferotemporal cortex (IT), whose connection with object codes are discussed in Chapter 5. Sparse coding has furthermore been applied in models of the early visual system, especially for models of simple and complex cells in the primary visual cortex [e.g., Olshausen and Field, 1996, Bell and Sejnowski, 1997, van Hateren and van der Schaaf, 1998], and many other areas [reviewed in Graham and

[6]Some implementations of ICA maximize squared cumulants such that solutions with strongly negative cumulants can occur that are highly non-sparse.

Field, 2007].

2.3.1 Sparse Codes in Neural Systems

The brain can be considered as a computational device that ultimately transforms sensory information into motor commands. Computation is an abstract process, but the brain is a physical system and thus information needs to be encoded by physical hardware, that is, by neurons. As there are extremely many conceivable coding schemes for neural hardware, this section introduces sparseness as a basic descriptive feature of such schemes. Two distinct, although often related, forms of sparseness can be distinguished. A unit is *temporally sparse* or exhibits *lifetime sparseness* if it is active at few time points only (i.e., it is inactive for the majority of stimuli and is selective for few stimuli). *Population sparseness*, on the other hand, is a property of a system where most population members are inactive at any given time point. If a system exhibits population sparseness, its units are sufficiently different (e.g., decorrelated) and stimuli are roughly equally likely, its units are also temporally sparse. Conversely, if a number of temporally sparse units are sufficiently different (e.g., decorrelated), they also show population sparseness. These two types of sparseness often occur together but are not identical [Willmore and Tolhurst, 2001]. A code, where only a small subset is active for any given stimulus and each stimulus is encoded by a different subset, is sometimes called *sparse-distributed* [e.g., Field, 1994] or *sparse-dispersed* [e.g., Willmore and Tolhurst, 2001].

For a binary code, we can define the *activity ratio* as the ratio of active units to inactive units and thus as a measure of population sparseness. Equivalently, the activity ratio can be defined as the ratio of active time points (or stimuli eliciting a response) to inactive time points (or stimuli eliciting no response) as a measure of temporal sparseness of a single unit. In the extreme case of a *local code*, only one unit out of all is active at any given time, or in an alternative formulation, each unit codes for only one specific stimulus or stimulus feature. Such units are sometimes referred to as'grandmother cells' [Gross, 2002, Thorpe, 2003]. The floor indicator in an elevator, for example, uses such a local code. The opposite extreme with very high activity ratio is *dense distributed coding*. Binary codes like ASCII are of this type [Földiák and Young, 2003]. Figure 2.5 shows examples of local, dense, and intermediate codes for four stimuli that consist of horizontal or vertical bars of red or blue color. The highly sparse local code (panel A) assigns a single unique unit to each of the four stimuli. The sparse semi-local code (panel B) encodes each feature locally and thus uses two units in order to code for each feature (i.e., color and orientation). The dense distributed code (panel C) needs only three units in order to encode all stimuli but each individual unit only conveys little information about a stimulus feature.

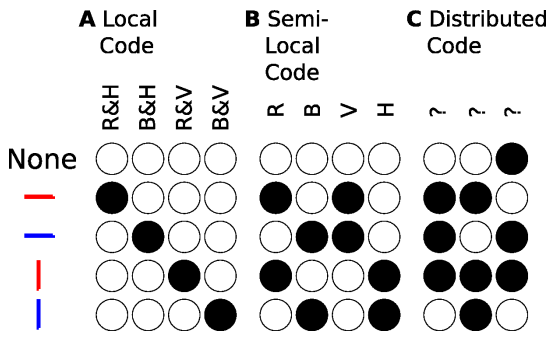

Figure 2.5: Illustration of local, semi-local, and distributed coding schemes. The stimulus is a bar of either horizontal or vertical orientation and either red or blue color. Filled circles indicate active units. A: In a local code, each stimulus is represented explicitly by a distinct unit. B: In the semilocal code, each feature of a stimulus is encoded locally: red (R), blue (B), horizontal (H), and vertical (V). C: In the distributed code, a dense but unintuitive combinatorial representation is used [adapted from Thorpe, 2003].

Sparse codes, as an intermediate approach between dense and local codes, have a number of advantages in the context of neural coding as they combine most advantages of local and dense codes but avoid most of their disadvantages [Hyvärinen et al., 2001, Olshausen and Field, 2004, 1997, Földiák and Young, 2003].
Compact codes require only few active components. Local codes are highly compact, but this feature comes at the cost of an extremely low representational capacity since N units can only encode N distinct entities. Sparse-dispersed codes on the other hand are also highly non-compact. Dense codes using the full combinatorial space of exponential size, which is 2^N for binary

codes, are maximally compact.
Storage efficiency in some recurrent memory systems is higher for population sparse codes than for dense codes due to reduced overlap of patterns [e.g., Willshaw et al., 1969, Baum et al., 1988] and higher than for local codes, which are limited by their extremely low representational capacity.
Highly local codes can be easier to decode (e.g., with linear readout units) than dense codes as there is little overlap of active units in different code words [Földiák and Young, 2003]. The downside of this property, however, is poor generalization performance, whereas sparse codes compromise between both properties. Local codes can also reduce wiring costs when downstream units require only few inputs as compared to dense codes.
Energy efficiency is often brought forward as a factor why sparse codes are advantageous for a neural system. In a system of spiking neurons, each spike consumes energy. If the goal of the system was to maximize information rate, half of the units would fire at each time point on average. However, such behavior would consume more energy than sparse firing. On the basis of metabolic energy constraints, Lennie [2003] shows that such behavior is impossible for the human brain. According to this publication, only as few as 1–10% of all cortical neurons in the human brain can on average be active concurrently, which implies high population sparsity and energy efficiency.
Sparse coding can reduce susceptibility to noise [Hyvärinen, 1999b, Willmore and Tolhurst, 2001] and increase robustness to unit failure as compared to local codes.
Local codes suffer less from the binding problem, because when two objects are perceived simultaneously, local codes simply have two active units whereas distributed codes are possibly undefined [Thorpe, 2003].
As a highly simplified summary, an encoding scheme has to be sufficiently dense to achieve adequate representational capacity but otherwise local enough to avoid the multiple drawbacks of dense codes. Since a brain can never experience as many stimuli in its lifetime as just ten thousand binary neurons could jointly encode in a fully exponential dense code, the representational capacity of a full distributed code might not be necessary anyway.

Sparseness Measures

For binary codes, the simple activity ratio, as introduced in the previous section, can serve as a measure of temporal sparseness. In non-binary codes, e.g., in a neuronal rate code, a unit can have many different activity levels for which the activity ratio is undefined. We can instead define a measure of temporal sparseness based on the activity histogram of a unit. For sparse units, this histogram should be strongly peaked around zero and the unit's seldom but significant activities cause "heavy tails" in the histogram. Many different sparseness measures based on the activity histograms are discussed

in the literature [e.g., Hyvärinen et al., 2001] but the most common index for sparseness is the normalized fourth moment or *kurtosis* of a variable:

$$\mathrm{kurt}(X) := E(X^4) - 3E(X^2)^2, \tag{2.8}$$

where

$$E(f(X)) = \int f(x) p_X(x) \mathrm{d}x \tag{2.9}$$

is the expectation value of the random variable X for a certain probability density p_X. High positive kurtosis values of a unimodal distribution indicate a strong "peakiness" of a distribution. As the standard Gaussian distribution has a zero normalized kurtosis, it is often used as a measure of "non-Gaussianity" in the ICA-literature [e.g., Field, 1994].
Another widespread measure of sparseness is given in Rolls et al. [1997b]:

$$a = \frac{\sum_{s=1}^{S} (r_s/S)^2}{\sum_{s=1}^{S} r_s^2/S} \tag{2.10}$$

with r_i denoting the firing rate for the i-th stimulus. This measure reaches its maximum of 1 for a unit that fires identically for all stimuli. For binary units, a value of 0.2 corresponds to a unit that fires for 20% of all stimuli and does not fire for all other.

2.3.2 Approaches for Sparse Coding

Sparseness describes a property of (neural) codes but not how to achieve such a code. A number of learning rules have been devised in the last thirty years that perform this task, the most relevant of which are introduced in this section.

Competitive Learning

Competitive Learning (CL) describes a family of learning rules that enforces units in a subsystem to compete for representing parts of possible codes. This competition during a training phase reduces the similarity of the representations of individual units by driving them apart in their input weight space. CL is often motivated as a model of the effect networks of laterally connected inhibitory interneurons have in the nervous system [Intrator, 2003].
Two basic forms of competitive learning can be distinguished. In *hard competitive learning*, for a given stimulus S the activation A of each unit U in a given layer is determined. The one unit U_{win} with highest activity A_{max} "wins the competition" to represent stimulus S and adapts its input

connections slightly towards this stimulus. The winning neuron U_{win} will therefore in the future be even more highly activated by a stimulus similar to S. In *soft competitive learning* on the other hand, many units can adapt their weights. Here, the amount of adaptation is usually a monotonic function of their relative activation strength. In this case, the unit with highest activation adapts most and the unit with lowest activation adapts least, if at all.
How should the weights of competing units be initialized before training? Like for neural networks, random values from Gaussian or uniform distributions can be used. Optimally these initial values should approximately cover the space of input patterns. Otherwise, especially for hard competitive learning, "dead units" can occur, i.e., units that never win the competition because their initial weight vectors are too dissimilar to any input pattern. Such units thus never learn and take no constructive part in the system. If the input data distribution is not known beforehand, units can be initialized with the values of training patterns in the beginning of the training phase. Hard competitive learning can suffer more strongly from the dependence on starting conditions (i.e., the weight initialization), as the highly local adaptation procedure is more prone to getting stuck in local optima.

Independent Component Analysis

Independent Component Analysis (ICA) denotes a class of algorithms for linear Blind Source Separation. Given a linear mixture of independent sources, ICA identifies a demixing matrix such that the outputs are "as statistically independent as possible" (cf. Figure 2.1). Independent codes minimize redundancy between outputs and have the advantage that the computation of the joint probability density of multiple simultaneously occurring events is simply the product of the independent probabilities [Olshausen, 2003]. Many different implementations for independence as a goal function have been proposed, including non-Gaussianity or temporal decorrelation across multiple time lags [reviews in Hyvärinen et al., 2001, Hyvärinen, 1999c].
The ICA algorithm applied in this thesis is called CuBICA and is based on the (cross-) cumulants of the output signals [Blaschke and Wiskott, 2004]. For independent data, offdiagonal elements of cross-cumulant tensors vanish [Hyvärinen et al., 2001, Blaschke, 2005]. Like most ICA algorithms, CuBICA assumes an initial whitening step as described in the beginning of this chapter. For whitened data, the first-order cumulant (i.e., the mean) vanishes and the second-order cumulants (i.e., the covariance) are already diagonalized. The remaining free parameters for rotations in the white subspace can be used to minimize the remaining higher-order cross-cumulants. [Blaschke and Wiskott, 2004, Blaschke, 2005]. The third-order cumulant (skewness) characterizes the amount of asymmetry and the fourth-order cumulant (kurtosis) the peakedness of a probability distribution. ICA algorithms based

on maximizing kurtosis thus directly maximize the most popular sparseness measure.

Other Approaches for Sparse Coding

Intuitively, maximizing higher order cumulants while keeping unit-variance and zero-mean constraints equates to maximizing peakedness of the outputs. This is because variance sums over the squared signal, whereas higher cumulants apply higher exponents to the signal before summing. More formally, according to [Hyvärinen et al., 2001, p. 374] many such functions of the form $E\{G(s)\}$ are estimators of kurtosis when G is a nonquadratic function, for example, $G(s) = -|s|$, and s is normalized to zero-mean and unit variance. Similarly, sparse representations can also be found by the simple operation of maximizing the output maximum under zero-mean and unit-variance constraints.

2.3.3 Application of Sparse Coding for the Unsupervised Learning of Place Cells

In a sparse code many components are inactive and only few components are significantly active at any given time. In the previous section it was argued that such a code has a number of advantages in a biological system. However, on the level of primary sensory signals, stimuli are typically encoded in a highly distributed non-sparse way. Sparse coding is one plausible way of transforming distributed codes into representations which are similar to those measured in the brain. This section is based on the article by Franzius et al. [2007b] and sketches the application of sparse coding for the transformation of a distributed code into a sparse code for a concrete example in the hippocampus: unsupervised learning of place cells. The biological background of the hippocampal formation is explained in more detail in Chapter 3. In summary, grid cells in entorhinal cortex fire in a very regular grid-like spatial structure (see Section 3.6). In contrast, place cells typically have none or only one single spatially localized firing field in a given environment (see Section 3.4). As it is likely that grid cells provide major input to the hippocampal regions containing place cell, the question arises how these localized representations could be formed on the basis of distributed grid cell input.

For this purpose, we simulated a fully connected linear two-layer network. The input units were 100 simulated grid cells of a virtual rat with activity patterns synthesized by Gaussians arranged on a hexagonal grid (Figure 2.6A). Some positional jitter, random anisotropy, and amplitude variation of the Gaussians was introduced and white noise was added to qualitatively match the slightly irregular experimental data.

Let $g_i(\vec{r})$ denote the activity of grid cell g_i as a function of location $\vec{r}$. Given

a virtual path $\vec{r}(t)$ of a rat within the enclosure, the input into the hippocampus coming from the grid cells is $x_i(t) := g_i(\vec{r}(t))$. To achieve sparseness we applied independent component analysis (ICA) [Hyvärinen, 1999c] on a set of 200.000 time points on the full set of 100 inputs by subtracting the mean and using the CuBICA algorithm, which attempts to diagonalize the tensors of third and fourth order cumulants [Blaschke and Wiskott, 2004], but we have obtained similar results with other sparsification algorithms, such as FastICA [Hyvärinen, 1999a] or simply maximizing peak activity under a unit variance, zero mean, and decorrelation constraint. The sign of each output unit, which is arbitrary for ICA, was chosen such that the value with the largest magnitude is positive, and then constants c_j were added to ensure nonnegative values. This yielded an affine transformation with matrix T producing 100 output signals $y_j(t) := \sum_i T_{ji} x_i(t) + c_j$ that are maximally independent and significantly sparser than the input signals (kurtosis increased on average from 2.8 for the input units to 27.3 for the output units). The output-unit activities as a function of location are $p_j(\vec{r}) := \sum_i T_{ji} g_i(\vec{r}) + c_j$ and show localized place fields (Figure 2.6G). We measured the number of peaks in a unit's output by counting the number of distinct contiguous areas containing pixels with at least 50% of the unit's maximum activity. A large proportion of output units (75%) show a single spot of activity (Figure 2.6G, units 1, 25, 50, 75), some units (6%) show few spots (Figure 2.6G, unit 79), both being consistent with the patterns of physiological place cells. Only few output units (19%) show patterns of activity without clear structure (Figure 2.6G, unit 100). The size of the resulting place fields is similar for most units and comparable to the size of the smallest grid cell fields, but it also depends on the number of grid cell inputs. More inputs lead to more localized output fields, while too few inputs can increase the number of fields per output unit (note that the number output units is always the same as the number of input units and the connectivity is complete).

There are different ways of achieving sparseness and localized place fields. We have used ICA here and have obtained similar results by maximizing peak activity. For a more biological plausible implementation, we use competitive learning (CL). The weights of the units are initialized with the firing rate of the grid cells at a particular location, with a different location for each unit. This is to avoid "dead units", i.e., units that never win the competition and thus never learn, but since in any given environment a significant proportion of place cells is inactive, a random initialization leading to some "dead units" might be considered realistic as well. In our case, the resulting code already is fairly sparse and localized (mean kurtosis: 9.9, number of units with single peak: 49, see Figure 2.6E). After competitive learning, kurtosis increases to 10.2 and the number of units with single peaks increases

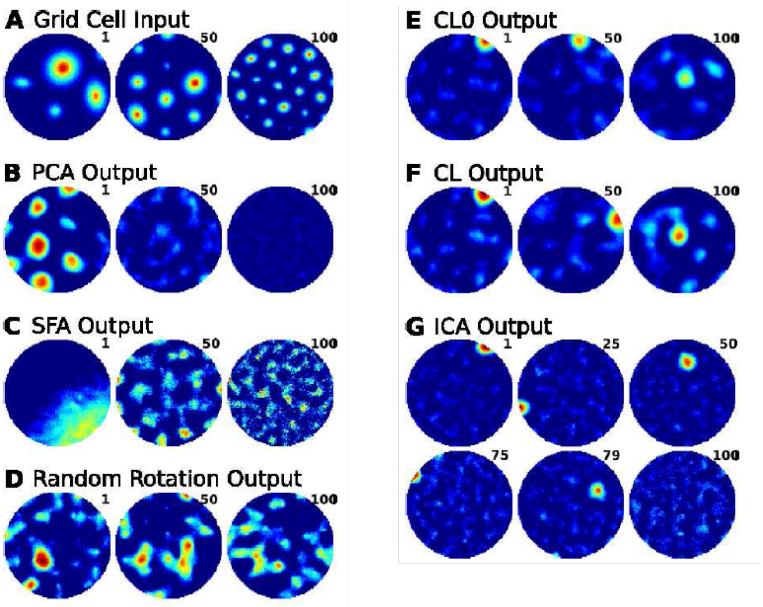

Figure 2.6: Inputs and outputs of different transformations for place cell generation from grid cells. A: Spatial firing pattern (SFP) of the input units representing grid cells. Three out of 100 units are shown. **B:** SFP of 1st, 50th, and 100th output computed by principal component analysis, ordered by eigenvalues. **C:** SFP of 1st, 50th and 100th output computed by Slow Feature Analysis, ordered by slowness. **D:** SFP of three out of 100 typical output units computed by random mixtures of the inputs. **E:** SFP of 1st, 50th and 100th output after initialization with sample vectors. Units are ordered by decreasing sparseness (kurtosis). **F:** 1st, 50th and 100th output after competitive learning, ordered by kurtosis. **G:** SFP of six out of 100 output units computed by independent component analysis as a means of sparsification, ordered by kurtosis. Place fields of sparser units tend to have higher peak activity and are more often located at the border of the enclosure, whereas less sparse units tend to have multiple place fields. Activities are color coded: red-high, green-medium, blue-zero activity. The full set of results can be viewed at http://itb.biologie.hu-berlin.de/~franzius/gridsToPlaces/

to 60 (Figure 2.6F). Furthermore, the output units are less correlated after competitive learning than before (mean absolute correlation drops from 0.189 to 0.014).

There are other linear transformations, however, that do not lead to localized place fields. As controls we have applied random mixtures, principal component analysis (PCA), and slow feature analysis [SFA; Wiskott and Sejnowski, 2002] to the grid cell input. The latter minimizes the mean squared time derivative of the outputs and has been chosen because Wyss et al. [2006] have presented a model based on the slowness principle that was able to learn localized place cells. As one would expect, with random rotations of the input the results retain some grid structure but are less regular than the input (Figure 2.6D) and no unit has one single or two peaks of activity. With PCA the first units (i.e., those with highest variance) are highly structured and have large amplitudes, much like the grid cells themselves, while the later low-variance units have low amplitudes and are noise-like (Figure 2.6B). None of these units had a single or two peaks of activity. From the temporal slowness objective we would expect patterns with low spatial frequencies first, and high-frequency non-localized patterns later, when outputs are sorted by slowness (Figure 2.6C). None of these outputs have only one or two peaks of activity. None of these three alternative linear transformations (Figure 2.6B–D) leads to localized place fields. Different starting conditions may lead to different results, but 5 out of 5 simulations showed the same qualitative behavior.

We conclude that sparse coding is a simple and efficient computational approach for the generation of place cells from grid cells. The mean kurtosis and percentage of localized place fields increase from 9.9 and 49% for the simple initialization with input vectors over 10.2 and 60% after competitive learning to 27.3 and 75% for the ICA algorithm, respectively. Other methods we have tested, such as random rotations, PCA, and SFA fail completely in generating localized place fields. The fact that SFA fails in our simulations is inconsistent with the results from Wyss et al. [2006]. Possibly, their model contains some hidden mechanisms that favor sparseness in addition to slowness. The simple initialization with input vectors is extremely quick and already fairly efficient [cf. McNaughton et al., 2006]. Such a simple mechanism might be a way for the almost instantaneous formation of place fields in a new environment. However, competitive learning still improves on that significantly while preserving many of the place fields chosen by the initialization process (Units 1 and 100 in Figure 2.6E–F maintained their place field while Unit 50 did not). Thus, competitive learning (or any other sparsification method) could be used as a refinement. ICA once again improves on the results of competitive learning but is biologically less plausible. There is some indication that grid cells reshuffle their phases if the animal is placed in a new environment (McNaughton, 2006). We have found that this results in output units like those with random rotations even if the place fields were

localized before the reshuffling. Thus, in our linear mode, for a successful remapping either the phases would have to change in some coherent way or the connectivity has to readapt. We believe the latter is more likely and we have seen above that it can be done rather quickly. However, even if sparseness is efficient in creating place fields from grid cells, the complexity of place field formation is now only shifted to the computation of grid cell behavior. A model for the formation of a distributed grid like spatial representation from quasi-natural sensory stimuli is presented in Chapter 4.
The use of CL and ICA for modeling place cells, head direction cells, and spatial view cells is described in more detail in Chapter 4.
Another well-known application of sparse coding comes from the area of object recognition and is implemented in the VisNet model [Rolls and Deco, 2002]. This hierarchical network employs feed-forward connections in vertical direction, and lateral competition within each layer by CL (see Chapter 5.3.1 for a more detailed discussion).

Chapter 3

Spatial Codes in the Brain

> *"Space plays a role in all our behavior. We live in it, move through it, explore it, defend it. We find it easy enough to point to bits of it [...] yet we find it extraordinarily difficult to come to grips with space. [...] Do we construct it from spaceless sensations or are we born with it? Of what use is it?"* O'Keefe and Nadel [1978]

This chapter gives a short overview of anatomical, physiological, and functional data relevant for spatial coding in the brain, together with a discussion of existing models. The literature on the hippocampus and on spatially correlated neurons in the brain is vast (a recent keyword search for "hippocampus" on pubmed returned more than 80,000 publications) and is growing rapidly. Additionally, the recently found grid cells in the entorhinal cortex (see Section 3.6) will most likely trigger an avalanche of new experimental and theoretical publications. Integrating only a fraction of all the available data into a computational model of spatial coding in the brain could take a lifetime alone. Thus, only a selection of the most important experimental findings relevant to this work is given below. Besides the cited original research papers, three books on the hippocampus and spatial codes in the brain are especially relevant for this work. One of these is the classic book by O'Keefe and Nadel [1978] *The Hippocampus as a Cognitive Map* from 1978, which is still highly readable despite its age. Although also somewhat outdated due to its publication before the finding of grid cells, Redish's book *Beyond the Cognitive Map* includes the most extensive reviews on experimental data available today [Redish, 1999], perhaps only rivaled by the recently published *Hippocampus Book* that will probably become a standard reference about the hippocampus [Andersen et al., 2007].

Our brains can directly sense stimuli from different modalities like pressure waves (with our ears), photons (with our eyes), or volatile chemicals (with our nose) and has specialized sensory areas for these tasks. We can,

however, not directly sense physical space – instead we have to rely on indirect clues from multiple sensory modalities including vision, audition, and touch in order to estimate our own position in space[1]. Accordingly, the mammalian brain has no specialized area for a sense of space like it has for example for auditory or visual sensory modalities [Fenton, 2007]. The brain structures involved in processing space lie deep in the brain many synapses away from any sensory receptor. Neural correlates of a spatial position were first found more than 35 years ago in the hippocampus of rats by O'Keefe and Dostrovsky [1971]. These *place cells* typically only fire when the animal is within a small contiguous area of the experimental arena. Much evidence has since been accumulated that no simple sensory stimulus triggers the firing of place cells (as it is possible for many neurons in sensory areas) but instead the best correlate of the cells' firing is the position of the animal in a given environment. Place cells are discussed in more detail in Section 3.4.

Correlates of head orientation were found twenty years later by Taube et al. [1990] in a neighboring area of the rat's hippocampal formation. These *head direction cells* fire maximally when the animal's head is turned into a specific direction (i.e., the cell's preferred direction) independently of the animal's position or behavior. This cell type is discussed in Section 3.5.
The latest addition to the list of neurons coding for certain aspects of space was recently identified in the entorhinal cortex (EC) of rats by Hafting et al. [2005]. These neurons show a regular hexagonal firing pattern, are accordingly named *grid cells* and discussed in Section 3.6.
Most data on place cells and head direction cells has been recorded from rodents and specifically from rats. Place cells and head direction cells were also found in primates, but in these animals yet another cell type that codes for another aspect of space was identified. *Spatial view cells* do not encode the animal's own (idiothetic) position (like place cells) but instead fire whenever the animal views a certain part of the environment [Rolls, 1999, 2006]. Properties of spatial view cells are discussed in Section 3.7.
A typical recording setup for hippocampal cells in rats is explained below in Section 3.2. The anatomy of the hippocampal formation is shortly explained in Section 3.3, as in this brain area the majority of oriospatial cells were found.
In the following, all these neurons coding for some aspect of spatial position, orientation or view will be summarized under the term of *oriospatial cells*. All oriospatial cells selectively encode some aspects of position and/or orientation of the animal, while being invariant to others. Figure 3.1 illustrates the differences between idealized examples of the different cell types. In reality, however, the picture is not so clear: place cells, for example, can strongly

[1]Some animals, including some migratory birds, possess a sense of the geomagnetic field, which is the closest equivalent to a direct sense of global orientation [Deutschlander et al., 1999].

depend on head direction in certain behavioral paradigms. Sections 3.4 to 3.7 summarize the invariance properties and other relevant experimental data of each oriospatial cell type along with a selection of theoretical models about the respective cells. In Section 3.8 the types of interactions and functional dependencies between the different oriospatial cell types, as far as they are known today, are summarized.
The information about its momentary own position and orientation in space provided by oriospatial cells can be used by an animal to purposefully change its spatial position. However, this process of *navigation* does not necessarily require self-localization. The different forms of navigation are summarized in the following section.

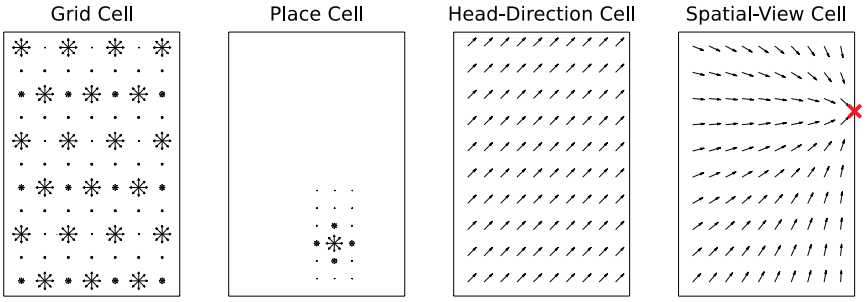

Figure 3.1: Spatial and orientation tuning of an idealized grid cell, place cell, head direction cell, and a spatial view cell. The activity of a grid cell is mostly orientation-invariant and not spatially localized but repeats in a hexagonal grid, whereas a place cell is also orientation-invariant but spatially localized. The activity of a head-direction cell shows a global direction preference but is spatially invariant, and the spatial view cell is maximally active when a specific view is fixated (indicated by 'x') with an amplitude that is independent of spatial position. Each vector denotes the activity of a cell at the vector's starting point with a magnitude corresponding to the vector's length if the animal heads into the vector's direction.

3.1 Self-Localization in Space as a Basis for Navigation

Navigation describes strategies to purposefully change the own position in space. This often requires the ability to identify one's own position and the position of the goal in space, in order to find a path between these positions. This ability is one of the most fundamental prerequisites for higher animals, as it is the basis for finding food and mating partners and for evading predators.

A thorough review of navigational strategies is beyond the scope of this thesis. But since self-localization in space is a major part of this work and a necessary prerequisite for complex navigational strategies, a short overview on navigational strategies is given below.
Redish [1999] defines five basic strategies of rodent navigation for the example of the Morris water maze task. In this task, a rat is placed in a container filled with an opaque liquid where a platform is hidden directly under the surface, invisible to the animal. Although rats are good swimmers, they are strongly motivated to find the hidden platform as a resting place in such an experiment.
If the animal has no information about the position of the platform, it performs a *random search*. A systematic search pattern might be more efficient but such behavior is usually not reported for rodents[2] [but note Wolfer and Lipp, 2000]. Random navigation is typically also used during exploration of new environments or if an animal is unable to determine its own position. This strategy requires no sensory stimuli and no knowledge about the own position in the environment, and it can be solved by a purely reactive system without any internal model of space at all. Besides allowing to find the hidden platform or a hidden food source by chance, random navigation can help the animal to acquire knowledge for more purposeful navigation later. This behavior is called latent learning and was shown first by Tolman [1948] with rats that explored a dry maze while they were neither hungry or thirsty. The rats were then deprived of food and water before they were released back into the maze, where the animals directly went to the food and water source they had previously seemingly ignored.
If a salient cue directly indicates the position of the submerged platform or if the platform itself is visible, the animal can directly swim towards it. This is called *taxon* or *beaconing navigation*. This type of navigation does not require the ability to locate oneself in the environment but only the abilities to identify the goal by the salient cue and to move towards it. If, however, the most direct path to the beacon is blocked (e.g., by a wire fence) this navigation strategy fails.
If the spatial relation between the animal's starting position and the platform is constant (i.e., reachable by a fixed path), the animal could use a fixed motor program to find the platform. Such a strategy is called *praxic navigation* and can, for example, be based on the sequence from prior random navigation. Given a sufficiently exact motor program, this strategy requires no sensory information and no self-localization to reach the goal.
A slightly more demanding, but still purely reactive, form of navigation is *route navigation*. This strategy is a combination of taxon and praxic naviga-

[2] If available, some rodents will use small objects as spatial markers and employ a better-than-random navigation strategy. Such markers are probably used for marking areas the animal already investigated [Stopka and Macdonald, 2003].

tion. Here, the animal uses a specific motor program when a specific sensory cue is perceived, e.g., it moves in a certain direction for a certain distance if it perceives a particular view. This approach can be characterized as a sequence of stimulus-response-stimulus commands as for example used in hiking guides [O'Keefe and Conway, 1978].
However, random, taxon, praxic, and route navigation are often inefficient or limited in their application in complex and dynamically changing environments. If the environment changes, a beacon is obscured, or if a route is to be reversed for the way back to the origin the latter three navigation strategies will likely fail. The most general and complex navigational approach is called *locale navigation*. Here, the direction to the goal is not indicated by a salient cue, nor is the route indicated by a stereotyped praxic movement sequence: instead, the animal has to build up an internal representation of space first. This internal representation is often termed a *cognitive map* of the environment, as first defined by Tolman [1932, 1948]. For successful locale navigation, the animal has to identify the position and orientation of the goal and of itself in the internal map representation. Based on this map, different routes can be planned (e.g., the quickest, the shortest, etc.). Some complex forms of rat behavior like actively finding shortcuts in a maze, can only be explained by locale navigation.
Random, taxon, and praxic navigation require only simple cue identification and no knowledge about the own position in space. For route navigation possibly more complex cue sets have to be identified for successful navigation. Still, the animal does not need to estimate its own position in space in this case - a simple stimulus-response function is sufficient. The most general case of locale navigation critically depends on self-localization relative to environmental cues.

3.2 Experimental Setup for Oriospatial Cell Recordings

The activity of oriospatial cells has been measured in thousands of experiments with electrophysiological recording techniques. In a typical experiment, extracellular local field potentials (LFP) are recorded in the brain of a freely behaving rat. Often a circular or quadratic arena of approximately 1 m diameter is used as a confinement for the animal during recording. The arena can either be open and allow the animal to see the surrounding laboratory or can be closed by high walls and/or a curtain, which allows a better control of stimuli perceivable by the rat. The latter configuration is especially important when the effects of cue manipulations (e.g., rotations of a prominent visual cue card) on the firing properties are tested. For complete cue control, the arena floor needs to be cleaned regularly because rats place urine marks, which they can use for novelty detection in navigational tasks

[Hopp and Timberlake, 1983, Genaro and Schmidek, 2000]. Sometimes a white noise source is added to the setup to occlude sounds from the laboratory which might serve as a directional cue for the animal [e.g., Sharp, 1996].
While the animal is in the arena, LFPs are recorded by means of chronically implanted microelectrodes in the hippocampal formation (see Section 3.3). Modern electrophysiological recordings typically employ microelectrodes with multiple blunt wires (tetrodes with four electrodes are the most common type today) as these allow a more efficient analysis of the local field potentials in order to assign the recorded wave forms to individual neurons by means of a "spike sorting" algorithm. Simultaneously to the LFP recording, the position of the animal is recorded with an overhead camera. In order to facilitate the automatic detection of the animal's position from the overhead video, the animals often carry one or more light emitting diodes (LEDs) on their head. If two ore more LEDs are used, also the head direction can be detected reliably[3], otherwise head direction is usually estimated as the derivative of current body movement. The latter approach is obviously restricted to episodes of animal locomotion and fails during episodes of pure head movement. Although rats can rear to considerable height as part of their search behavior (more than a quarter of the arena diameter for the typical 76 cm arena), only in very few experiments the three-dimensional position of the animal is recorded [e.g., Calton and Taube, 2005]. Usually only the two-dimensional position in the plane is measured.
Given the momentary position from the overhead video and the firing rate of a cell, a *spatial firing map* can be constructed. Newer publications based on experiments using tetrodes tend to show smaller place field sizes and fewer place cells with multiple firing fields than earlier ones that are based on recordings from single electrodes. This is probably due to improved spike sorting with tetrode data [Redish, 1999], as the spike sorting step may split one cell's spikes into multiple clusters or pool over multiple cells' firing. A further reason might be that a different cell type (likely theta cells, see Section 3.3) has sometimes been recorded together with complex spike cells. Tetrode recordings also have the advantage that they allow simultaneous recordings from multiple cells [theoretically up to 1,000 cells: Buzsáki, 2004]). All extracellular recording techniques have the disadvantage that they depend on the spike sorting algorithm. A very recent technique [Lee et al., 2006] allows intracellular recordings of freely moving rats. Such recordings are still very difficult to perform but allow an unambiguous identification of single cell activity and moreover the recording of subthreshold membrane potentials. The intrinsic disadvantage of intracellular recordings is the limi-

[3]For the case of two LEDs a difference in color or brightness allows to unambiguously determine the heading of the animal. Otherwise the movement direction estimated from prior video frames can serve to identify the animal's head and tail as the animals usually do not move backwards much.

tation to a single cell per electrode and the resulting impossibility to record from large populations of neurons simultaneously. Furthermore, the maximal recording duration of this technique is shorter than that with extracellular electrodes.
Usually the animal is encouraged to move much within the recording arena in order to smoothly sample the spatial firing map. In a widely employed experimental setup a rat searches for food pellets that are dropped into the arena at random positions in regular intervals. Since the animal typically is mildly food-deprived, this setup encourages the animal to constantly move throughout the arena in search of food. As a result, the animal performs seemingly random movements through the arena for several minutes (c.f. Section 3.1). With this paradigm different behaviors (searching food, consuming food, grooming, sniffing, etc.) are distributed rather homogeneously over the arena[4]. This means that any nonspatial correlates of the recorded cells are likely to be averaged out in the firing map.
Recordings in the open field require long sessions if a complete firing map is to be recorded, since the animal has to visit each position at least once, but preferably many times, while heading into different directions. This sampling problem is alleviated if instead of an open field an essentially one-dimensional arena is used. Linear and circular tracks are environments of this kind: here the animal typically moves in one direction, only turning at track ends of the linear track. Radial mazes (e.g., plus- or 8-arm mazes) are the third popular arena type and can be considered as a combination of linear tracks, albeit in some configurations the center area is so large that this part is more similar to an open field.
In summary, the basic experimental setup has stayed remarkably stable over the last 30 years. Despite advances with the recording equipment, only very recently larger arena sizes have been used, which allowed the finding of grid cells (see Section 3.6). Other new approaches include the use of a virtual reality setup [e.g., Hölscher et al., 2005], which allows arbitrarily large virtual environments and possibly also better cue control.

3.3 The Hippocampal Formation

This section sketches the most basic facts about the hippocampal formation necessary as a context for the data later in this chapter. For in-depth discussions of hippocampal neuroanatomy and physiology see, for example, Amaral and Witter [1995], Greenstein and Greenstein [2000], Andersen et al. [2007] and references therein.

[4]Nevertheless, rats have a tendency to stay close to the arena walls.

3.3.1 Anatomy

The hippocampal formation is part of the limbic system, which also contains the limbic lobe, the amygdaloid nucleus, and the anterior nucleus of the thalamus, and is located centrally in the brain in the temporal horn of the lateral ventricle under the temporal lobe [Greenstein and Greenstein, 2000]. The hippocampus is one of the most extensively studied brain areas and has been investigated for centuries [Andersen et al., 2007]. This long history of research has led to some inconsistencies in the terminology. According to the terminology by Amaral and Lavenex [2007], the *hippocampus*[5] *proper* is divided in the three subfields CA1, CA2, and CA3. CA stands for *Cornu Ammonis*[6], which is an alternative term for hippocampus. Some authors only differentiate between CA1 and CA3 while others describe CA4 as a fourth area, but the majority of experiments are carried out in the unambiguously labeled areas CA1 and CA3. The hippocampus proper is part of the larger *hippocampal formation* that also includes the dentate gyrus (DG, also called Fascia Dentata), the entorhinal cortex (EC), and the subicular complex. The subicular complex contains the subiculum, parasubiculum, and postsubiculum (which is also known as the (dorsal) presubiculum).
Rapp and Gallagher [1996] estimate the number of neurons in one rat hippocampus to be 1.2 million granule cells in DG, 225,000 neurons in CA3/2 and 390,000 neurons in CA1. Since there is one hippocampus in each brain hemisphere, the total cell numbers for the rat brain are twice the given sizes. Combining these estimated cell numbers with the results from Olbrich and Braak [1985][7], a rat CA3/2 contains approximately 212,000 pyramidal cells and CA1 contains circa 367,000 pyramidal cells.

The hippocampus differs from most brain regions by its simpler structure [Burgess and O'Keefe, 2003] and because it is "one of the few brain regions that receives highly processed, multimodal sensory information from a variety of neocortical sources" [Amaral and Lavenex, 2007, Burgess and O'Keefe, 2003], including vision, olfaction and audition. Another striking difference to the neocortex is that the hippocampus has much less reciprocal connections, i.e., if hippocampal region A projects to region B, region B typically does not project strongly back to A. Instead, the hippocampal formation forms a large loop, called the *trisynaptic circuit* that begins with the EC where most cortical input to and output from the hippocampus and

[5] A Hippocampus is a sea horse, which has a similar curled form and size as the human hippocampus.

[6] Cornu Ammonis means "Ammon's horn" after the Egyptian god Amun Kneph whose symbol was a ram [Amaral and Lavenex, 2007]. The ram's horn resembles the hippocampus.

[7] Olbrich and Braak [1985] report 9.4 % ± 1 % nonpyramidal cells in hippocampus proper. These results are based on human hippocampi but similar ratios of 10 %–20 % are cited for rat and monkey in this article.

subicular areas passes through. The circuit continues to the granule cells in the DG (via the "perforant path"), and from DG to CA3 (via the "mossy fibers"). CA3 has massive recurrent connections projecting onto itself but also projects to CA1 (via the "Schaffer collaterals"), and the circuit finally closes the loop by projections from CA1 back to the EC [Nakazawa et al., 2004]. CA1 also provides the major input for the subiculum, which in turn has major projections to the EC but does not project back to CA1. The notion of the trisynaptic circuit EC/DG/CA3/CA1 is, however, highly simplified as there are many more connections between hippocampal areas, e.g., direct connections of the perforant path from EC to CA3. The trisynaptic loop as the basic hippocampal architecture is similar in rodents and primates, including humans, although there are some substantial differences in the connectivity, layer structure, and size [Amaral and Lavenex, 2007].

3.3.2 Cell Types

Ranck Jr. [1973] coined the term *complex spike cells* for the most prominent cell type observed during local field potential recordings in the hippocampus. The name of this cell type is derived from its firing properties as it sometimes fires in bursts of action potentials with decreasing amplitude [Muller et al., 1987]. As complex spike cells are almost always place cells and all hippocampal place cells seem to be complex spike cells, the terms "complex spike cell" and "place cell" will be used synonymously in the following [Fox and Ranck Jr., 1981].
The main excitatory cell type in dentate gyrus is the granule cell with a proportion of circa 88 % of all neurons in the area, whereas the other cells seem to be inhibitory theta cells [Jung and McNaughton, 1993]. Like in hippocampus proper, theta cells show higher average firing rates (around 8 Hz), low spatial selectivity, and a strong theta modulation. Granule cells exhibit low average firing rates (around 0.2 Hz) and most cells have one or more place fields. Place fields of granule cells appear less homogeneous than in CA, as some granule cells have a single continuous place field while others have many small discrete place fields [Jung and McNaughton, 1993].

3.3.3 Functional Role of Hippocampus

The hippocampus is strongly involved in memory and spatial processing. The debate whether only one of the two or both are the main function of hippocampus has been going on for many years [e.g., Eichenbaum et al., 1999, O'Keefe, 1999, Redish, 2001]. The remarkable spatial correlates of cells in the hippocampal formation are discussed in the following sections. The most prominent example of hippocampal memory function comes from a patient known as H.M. whose medial temporal lobes were partially removed for treatment of severe epilepsy. Patients like H.M. with damaged temporal

lobes often suffer from anterograde amnesia, that is, the inability to form stable memories of episodes after the brain damage, but have unaffected memory of episodes much older than the incident. Time periods directly before the incident are gradually affected by the retrograde amnesia. The amnesia only affects declarative or explicit memories and does not affect implicit memories or procedural skills that are acquired over multiple sessions like mirror reading or certain motor tasks. The basic theory of hippocampal involvement in memory is that new declarative or explicit memories are first stored as short-term memories in highly plastic hippocampal synapses. These hippocampal memories are then used to entrain (consolidate) long-term memories in the neocortex [McClelland et al., 1995] and that after consolidation long-term memory recall is independent of the hippocampus. In the context of spatial memories some evidence suggests that the EC might be necessary for recent spatial memory and the neocortex for remote spatial memory [Steffenach et al., 2005]. The vast literature about the function of hippocampus for (nonspatial) memory will not be reviewed here. The following sections discuss nonspatial influences on the different oriospatial cell types in hippocampus in the context of spatial tasks. It is possible that the spatial information is already available explicitly upstream in the EC (cf. Section 3.6) and the hippocampus mainly recodes spatial information within a more general framework of memory formation. This hypothesis is supported by the model in Chapter 4.

3.4 Place Cells

A place cell can be defined functionally as a neuron whose firing is strongly correlated with the spatial position of the animal. More specifically, a place cell fires only when the animal's head is within one (or few) small contiguous areas of a given environment. Figure 3.1 on page 27 illustrates the spatial firing of an idealized place cell and Figure 3.2 shows a recording of 80 hippocampal cells (including pyramidal and inhibitory cells). Alternatively, a place cell can be defined anatomically as a pyramidal (or complex spike) cell in hippocampus proper[8]. Some authors include dentate granule cells [Jung and McNaughton, 1993] under the term place cell as well or even more generally all neurons with spatial correlates in other areas. Such areas include the basal ganglia, primary sensory motor cortex, medial entorhinal cortex, subiculum, parasubiculum, frontal cortex, and lateral septum [Knierim, 2006].

[8] A complex spike cell in hippocampus without a place field is sometimes called "a quiet place cell", especially if the same cell has a place field in a different environment.

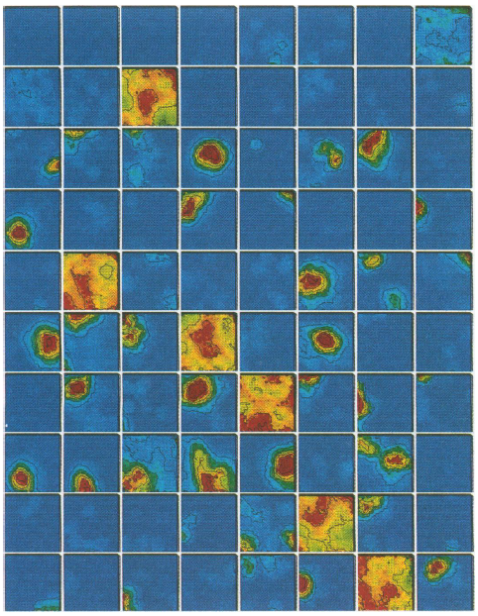

Figure 3.2: Firing fields of 80 simultaneously recorded cell in the rat hippocampus. Recordings from both pyramidal and inhibitory neurons in the hippocampal formation of a rat during unrestrained exploration in a familiar environment are shown. Each panel represents the spatial distribution of the firing rate for one cell. Maximal rates for cells with significant spatially related firing are indicated by red; no firing is indicated by dark blue. The inhibitory cells exhibit more dispersed firing. Reprinted from [Wilson and McNaughton, 1993] with permission.

3.4.1 Spatial and Nonspatial Determinants of Place Cell Firing

Place cells were discovered in 1971 by J. O'Keefe and J. Dostrovsky [O'Keefe and Dostrovsky, 1971] in the hippocampal area CA1 of the awake and freely behaving rat. While early results by Ranck characterized complex spike cells as both correlated with behavioral and spatial data, the highly influential book by O'Keefe and Nadel, *The Hippocampus as a cognitive map*, from 1978 supported the perception of hippocampus as an area coding almost exclusively for space. This view has subsequently been opposed by findings of other nonspatial correlates of place cell firing. These findings, many of which are reviewed in [Redish, 1999], show correlations of place cell firing with running speed, running direction, and turning angle [McNaughton et al., 1983a, Wiener et al., 1989], type of texture on the floor under the animal [Young et al., 1994], odor [Eichenbaum et al., 1987, Eichenbaum and Cohen, 1988, Wood et al., 1999], arousal [O'Keefe, 1999], behavioral task [Markus et al., 1995, Wood et al., 1999], and stage of task [Eichenbaum et al., 1987, Otto

and Eichenbaum, 1992, Wood et al., 1999, Frank et al., 2000, Wood et al., 2000]. The correlation with running speed is especially evident as place cells cease firing when the animal is immobile, and even when it is passively moved but tightly restrained [Foster et al., 1989]. Czurkó et al. [1999] have found a strong impact of running speed on place cell firing rates even in a running wheel where firing rates depended on running direction and position of the running wheel and increased linearly with running speed.
Even if a certain cell in the hippocampus exclusively codes for spatial information, spurious correlations with nonspatial variables like stage of task are to be expected unless the experiments take special care to eliminate nonspatial influences. O'Keefe [1999] argues that most nonspatial correlates of hippocampal activity are spurious. For example, consider a place cell with a directional place field next to a task-specific goal area. Unless the goal area is shifted to different locations during the experiment, the place cell firing correlates with "goal approach".
The debate about the role of hippocampus in nonspatial tasks is ongoing. While the important role of the hippocampal formation for spatial processing is undisputed, no such clearcut and prevalent effect like place cell firing has been found yet for nonspatial processing.
Growing evidence suggests that the hippocampus and the entorhinal cortex projecting to it are functionally differentiated along the longitudinal axis, which is sometimes also called the dorsoventral or septotemporal axis [Jung et al., 1994, Moser and Moser, 1998, Hargreaves et al., 2005]. A higher proportion of complex spike cells in the dorsal part of hippocampus have place fields than those in the ventral part, and spatial selectivity of place cells in the dorsal part is higher than in the ventral hippocampus. Lesion studies show that small parts of the dorsal (or medial) hippocampus are sufficient for spatial learning but equally small parts of the ventral hippocampus are insufficient [Moser and Moser, 1998]. These findings suggest that the entire hippocampus of the rat might be involved in spatial coding but especially the dorsal part. The role of the ventral hippocampus besides spatial coding is still discussed – Hargreaves et al. [2005] suggest it could be involved in "item memory" (i.e., "what" information) complementary to the "spatial memory" (or "where" information) of the dorsal hippocampus. An alternative view is that place fields in the ventral hippocampus might be much larger and thus typically not detectable in small arenas [personal communication with A. Treves, 2007].

Place cells discussed so far were located in rat CA1 or CA3. The very different structure – CA3 has massively more feedback connections than CA1 – of the two areas strongly suggests some functional difference, although the firing fields of the place cells are usually reported as largely indistinguishable between CA1 and CA3 [Muller and Kubie, 1987, Markus et al., 1995, but note Yoganarasimha et al. 2006]. For a review of other non-hippocampal

place cells see [Redish, 1999, p. 274f]. Note that the cells in the medial entorhinal cortex described by Quirk et al. [1992] as larger and noisier place cells might be grid cells (see Section 3.6).

For the monkey hippocampal formation, most early data is only available for fixated animals during presentation of stimuli around the animals. In such recordings, object- and position-specificity of hippocampal neurons have been tested. For example, Tamura et al. [1992] report that 10 % of the recorded neurons show specificity for the relative location of the stimulus during the presentation of visual and auditory stimuli to a fixated animal. Later experiments with monkeys fixated within mobile chairs or carts provided closer similarity between monkey experiments and rodent experiments [e.g., Ono et al., 1993]. Such experiments by Ono and colleages found about 5 %–15 % of spatially selective hippocampal neurons [O'Keefe, 1999]. In order to overcome the limited mobility of monkeys in such recordings, experiments were performed in virtual reality environments by Hori et al. [2005], where 32 % of the recorded neurons in the hippocampal formation showed spatial firing correlates. In this study, a rearrangement of distal (virtual) cues caused a rearrangement of the firing fields of most cells, similar to the remapping phenomenon in rodents (cf. Section 3.4.5). Place cells in primate hippocampus might not be as prevalent as in rodents, or their function might depend on active exploration of the environment [Ludvig et al., 2004] as not all groups looking for place cells in primate hippocampus succeeded in finding them [O'Mara et al., 1994] when experimenting with passively displaced macaques. The finding of spatial view cells by Rolls and colleages are discussed below in Section 3.7. These results seem to indicate a dependence of hippocampal firing patterns on active exploration and possibly also on the concrete movement pattern of the animal, which is the main prediction of the model in Chapter 4.

Place cells were not only recorded in rodents and monkeys, but also in rabbits, dogs, cats, guinea pigs [Robinson, 1980], humans [Ekstrom et al., 2003], and recently Uanovsky and Moss [2007] recorded place cells in freely moving bats[9]. Place cells and spatial view cells (see Section 3.7 below) were also identified in the human hippocampus by Ekstrom et al. [2003] during the rare opportunities to record from humans that undergo invasive monitoring for surgical epilepsy treatment. In summary, place cells seem to be common in the mammalian hippocampal system and to encode spatial as well as nonspatial features.

3.4.2 Field Size, Number of Subfields, Field Distribution

A place cell in the hippocampus typically has high firing rates [of up to 500 Hz within bursts according to McNaughton et al., 1983b] within its place

[9] The bats were not flying but restricted to walking movements on a tiltable surface.

field and practically zero firing rates (below 1 Hz) outside the place field[10]. But as firing rates outside the place field are not exactly zero, for exact measurements of place field boundaries some firing rate threshold has to be defined. This threshold is typically defined to be "the background firing rate" of approximately 1 Hz or a rate some standard deviations above mean firing rate (often 2 SD). An alternative approach for estimating a spatial scale of place fields is to calculate the mean distance after which the correlation of two population vectors falls under a certain threshold [Maurer et al., 2005]. An idealized place cell (as depicted in Figure 3.1 on page 27) has exactly one Gaussian shaped place field in a given environment, but many variations are reported from electrophysiological recordings. Some place fields are crescent-shaped along the walls of a circular arena or have multiple separate firing fields [Muller et al., 1987, O'Keefe and Conway, 1978, McNaughton et al., 1983a], although the number of firing fields reported decreased with the advent of newer measuring techniques [(cf. Section 3.2), Redish, 1999].
When recording in different environments, a place cell is typically not active in all environments, although once a place field is established in a given environment it typically is stable over multiple sessions. Nakazawa et al. [2004] report that on average 30 % to 50 % of all place cells are active in any given environment. An earlier article reports only 12 % active place cells in a given environment [Thompson and Best, 1989]. These numbers may be influenced by the fact that a "quiet" cell is hard to identify during the electrophysiological recording and by an experimental bias to record from cells with high firing rates.

Most of the experimentally measured field size data are given as a fraction of the arena area, which typically is under 1 m^2. Does the field size of place fields change for larger or smaller arena sizes? Is the average place field size determined by the absolute size of the arena? Few experiments investigate this question, probably mainly because of problems in larger arenas due to increased sampling time and constraints of the electrophysiological setup with limited cable length. In the worst case, the results obtained from small recording chambers might be completely unrelated to behaviorally relevant spatial scales, since mice in the wild forage and navigate in areas of more than 10,000 m^2 [Benhamou, 1990]. The spatial size restrictions for electrophysiological recordings might be overcome in the future with the technique of Hölscher et al. [2005] who demonstrated that rats can navigate in a virtual reality environment where the simulated arena is not restricted in size. Field size varies along the long axis of the hippocampus. Jung et al. [1994] find that place fields of cells in the middle of the hippocampus are of twice the size (average 1.6 % of the apparatus area or 462 cm^2) as in the septal

[10] Maxima of most published firing rate maps are not above 10–20 Hz, though [e.g., Wills et al., 2005]. This might be due to averaging over multiple transitions through the place field with differing head directions or nonspatial variables.

part (average 6.6 % of the apparatus area or 1,874 cm^2). Muller reports an average field size of 13 % (590 cm^2) of the apparatus area of 4,537 cm^2 (and a range between 3 % (136 cm^2) and 50 % (2,268 cm^2) for hippocampal pyramidal cells [Muller, 1996]. As an alternative measure, Maurer et al. [2005] measure the spatial correlation length of all place field activities. Calculating the average distance over which the correlation between the firing rate population vectors fall to a value of 0.5 yields values between 25 cm in dorsal and 42 cm in middle hippocampus for the linear track independently of track length.

O'Keefe and Burgess [1996] investigated the effect of changing the side lengths of a rectangular arena between 61 cm and 122 cm on the firing properties of hippocampal place cells. Their results suggest that the field size of an already established place cell partially depends on the arena size. When one or two sides of the arena are extended some cells stretch in the elongated direction or even split up in two parts.

The population of all place fields in hippocampus has a strong overlap, which becomes evident if we combine the estimated figures from above. The estimated field sizes (average above 100 cm^2), a given apparatus size (below 1 m^2), the total number of neurons in one hippocampus (above 500,000), the percentage of active neurons in a given environment (above 30 %), and a homogeneous place field distribution results in a lower bound of 1500 place cells active in one hippocampus at any given position[11]. This estimate is certainly very rough but interesting to relate to the calculation of spatial information of coactive place cells by Wilson and McNaughton [1993]. The joint firing rates of approximately 130 CA1 place cells during one second are sufficient to estimate the position of the animal with an error below 1 cm, or about 380 cells to reach the same acuracy within 0.1 s.

Place fields in the center of the arena have a slight tendency to be larger than towards the periphery [O'Keefe, 2007], but generally place field centers tend to be equally distributed over the apparatus area and specifically no correlation between average dwelling time and the distribution of field centers exists [Muller, 1996] (but note [Hollup et al., 2001]). Other authors report some cases of increased place field frequency at the edges of the arena [O'Keefe, 2007].

In summary, a given position in a fixed environment is jointly encoded by a large population of hippocampal place cells.

[11] The number of coactive place cells in different areas of the rat brain can correspondingly be estimated to be 3600 in dentate gyrus, 640 in CA3/2, and 1100 in CA1. Under certain conditions, however, the number of active granule cells in dentate gyrus can be only 2 %–3 % [Chawla et al., 2005], thus reducing the estimate for dentate gyrus to 240–360.

3.4.3 Head Direction Dependence

Most place cells recorded in open fields are invariant to head direction but selective for the animal's position. Interestingly, the degree of orientation-invariance depends on the structure of the environment and on the animal behavior. On the linear track, most place fields are clearly directional.
Undirectional place fields may appear directional in measurements, especially if sampling time is short or the rat has certain biases during movement [e.g., Muller et al., 1994, Markus et al., 1995, Burgess et al., 2005]. One case of such spurious directionality can occur if the animal traverses more often through the center of the place field from one direction and through the border (or not at all) of the field in the other direction. This effect can be reduced if only traversals near the center of the place fields are included into the computation of directionality. A second source of spurious place field directionality is that the rat, due to its physical extent, cannot traverse a place field in all directions at the border of the arena. The directionality of place cells near the arena borders can thus not be determined exactly. Thirdly, spurious directional tuning can arise if the animal traverses with different speed through a place field since firing rates of place cells often depend on running speed[12]. Two approaches to overcome the problem of spurious directionality from biased sampling are discussed in Cacucci et al. [2004].
The behavioral and environmental determinants of place cell directionality in rats have been thoroughly tested experimentally in the seminal work by Markus et al. [1995]. The authors found no significant effect of environmental complexity on place field directionality, i.e., either a black curtain or many distinct objects surrounding the recording chamber had no significant effect. Less than 20 % of all recorded place cells were directional in a high-walled cylinder during the standard random pellet search task and less than 1/3 of the cells were directional on a circular open platform during the same task. Significantly more place cells were directional (ca. 2/3) in a radial maze (see Section 3.2) during a forced choice search task. The behavioral task of the rat (and thus the movement pattern) strongly influences directionality: after 30 minutes of random pellet searching, four equidistant points on the border of the circular platform were repeatedly and sequentially baited (first clockwise, then counter-clockwise). The mean directionality of place cells increased significantly during these linear movements in the second phase of the experiment. Directionality in the plus maze was generally much higher than on the circular platform but also increased in magnitude from the random to the directed search task. There was no significant difference in directionality in two different plus mazes with wide or narrow arms. Rats did not turn more often in wide arms, so that the actual (turning) behavior and

[12] Note that a positive linear dependency of firing rate with running speed would *prevent* spurious directionality.

not the ability to turn seems to determine the directionality of place cells. The authors conclude that the *"visual environment per se is less important a determinant of directional tuning than constraints (behavioral and environmental) on the animal's behavior"* and that *"hippocampal place fields are more directionally dependent when the animal is planning and/or following a specific route than when it is engaged in quasirandom foraging, involving erratic changes in the distance and direction of motion."*

The sampling problem of the three-dimensional joint position and head direction space in the open field (c.f. Section 3.2) could be alleviated by switching to a linear (or circular) track where the animal's spatial configuration space is basically only two-dimensional (one real-valued variable for the position and one binary variable for the direction). However, linear track experiments cannot easily be used as a lower-dimensional alternative for open field experiments, as most place cells behave differently in the linear track than in the open field. While the majority of place cells are head direction invariant in the open field, most are orientation-specific in the linear track or in the arms of a radial maze, i.e., the place cell only fires on some part of the track when the animal moves "north" but it does not fire at the same position when the animal moves "south". Some place cells fire both in "north" and "south" direction but at different positions on the track [Markus et al., 1995, McNaughton et al., 1983a]. O'Keefe and Recce [1993] report 14 out of 15 place cells in hippocampus proper that have place fields in a linear track of 1.5 m length only when the rat runs in one direction. One out of 15 cells had a place field for both directions but at different positions. In summary, place field directionality seems to be mostly influenced by the animal's behavior and especially by the linearity of movement. The more curved the animal's trajectory is in a given arena and the more directions are sampled, the less directional are its place cells.

3.4.4 Development of Place Cells in New Environments, Reliability and Stability of Place Fields

Do place fields exist the first time an animal enters a new environment or does it take time to learn them? The answer to this question is hard to find since the sampling of place fields in the open field takes many minutes and different experimenters report quite contradictory results. Furthermore, many articles do not report if the animal was familiar with the recording arena or laboratory room before the experiment.
Hill [1978] reports strong firing of place cells at the first passing of an animal through the area where the cell later displays a stable firing field, whereas Wilson and McNaughton [1993] report instable place fields during the first 10 minutes in a new environment, which stabilized during the next 10 minutes. Tanila et al. [1997] report that some cells already show strong place fields after few minutes, while others built up their field during

more than 30 minutes.
Even after more than 30 minutes when a place field is "well established", the firing of an individual place cell can be highly variable: a cell might fire maximally once during one traversal of its field and not at all during the next traversal [Muller, 1996]. This variability could simply be a noisy spatial representation or an unidentified additional variable encoded by the cell. Nevertheless, the firing maps obtained by averaging during single trials tend to be stable for as long as the recording electrode is stable, which can be over multiple months [Muller, 1996].
Steffenach et al. [2005] report that the recall of previously established place fields remains stable after lesions in the dorsocaudal region of the entorhinal cortex (this is the area where grid cells have been reported; see below) but that spatial learning in new environments is impaired after the lesion.
Frank et al. [2004] investigated place field properties over multiple days of exposure. The authors describe strong plasticity of place cells in the first minutes of exposure to a new environment but even larger changes on the second day of exposure if exposure on the first day was short. The authors conclude that a minimum of 5–6 minutes of exposure to a novel environment is necessary to form stable representations.

3.4.5 Environmental Manipulations

Many publications examine the effect of environmental manipulations on place cell firing. Although no simple stimulus seems to trigger place cell activity, even small environmental manipulations can have large effect on the firing pattern. The firing pattern of a population of place cells in one environment is in most cases orthogonal to (or statistically independent of) their spatial firing patterns in a sufficiently different environment [McNaughton et al., 2006, Quirk et al., 1990, Gothard et al., 1996]. If spatial cues (e.g., relative position of cue cards, shape of a flexible arena wall) are slightly changed, some place cells change their firing pattern in a (yet) unpredictable way. This effect is called *partial remapping*. If the relative positions of cues are altered more drastically or if cues are added or removed, a *complete remapping* of place fields can occur. After a complete remapping all[13] place cells change their activity patterns and some previously active cells become silent while other previously quiet place cells become active.

Leutgeb et al. [2005] distinguish "rate remapping" and "global remapping" as two subtypes of the remapping phenomenon in recordings from CA1 and CA3. In the case of rate remapping place cells keep their field locations but change their firing rates, whereas in complete remapping both positions and firing rates change to a statistically independent (orthogonal)

[13] A small percentage of cells keep their place fields, as expected to happen by chance.

representation. The two remapping types corresponded to two experimental paradigms of this publication. Firstly, in the "variable cue, constant place" paradigm, the recording took place in the same laboratory room with visible distal cues in the room but the recording arena was varied either in color or in shape. Secondly, in the "variable place, constant cue" paradigm, the recording arena remained constant but recording took place in two visually distinct laboratory rooms. As expected from earlier results, place cells tended to bind to distal cues in the case of cue conflicts in experiments with the "variable cue, constant place" paradigm. Here, firing patterns remained similar, so that *"without closer analysis, one might have concluded that there had been no substantial effect of cue-condition or environmental shape on the hippocampal code"*. Specifically, correlations between recordings with different cues remained high (median place field correlation of > 0.8 in CA3), median center of mass changes remained low ($3.3 - 9.8\,\text{cm}$ in CA3) but the median absolute firing rate changes where significantly higher than in control experiments. In contrast, experiments with the "variable place, constant cue" paradigm caused the population vectors to decorrelate (median place field correlation of -0.10) and the centers of mass of the firing fields shifted by a median of $34.5\,\text{cm}$ in CA3. The authors argue that two independent coding principles are used in hippocampus, one coding for spatial position of the animal in a given environment and one for the configuration of an individual environment. While rate remapping might account for some of the earlier remapping phenomena, there is also contradictory evidence that complete remapping can take place even when the spatial environment is unchanged. Further research seems necessary to investigate if the reported phenomena generalize to other experimental conditions.

If all visual cues are rotated consistently, place fields usually rotate with the cues in unison (coherently) [Rotenberg and Muller, 1997] without remapping. If cues are rotated in the presence of the animal, cue control over place fields is reduced [Poucet et al., 2003]. These findings demonstrate the strong influence of stable sensory cues on place cell firing. There is, however, also clear evidence of a strong influence of idiothetic cues on place cell firing [Moser and Paulsen, 2001, McNaughton et al., 2006]. Firstly, if significant landmarks are removed [O'Keefe and Speakman, 1987] or the light is switched off [Quirk et al., 1990], many place cells keep their stable firing fields. Secondly, most place cells in rats are non-symmetrical in symmetrical environments [Sharp et al., 1990]. Thirdly, firing patterns in two visually identical rooms are different if the animal is not disoriented during transport [Skaggs and McNaughton, 1998, Tanila, 1999], but this effect might also be due to other non-visual sensory cues .

Path integration or *dead reckoning* is a process of integrating self-motion over time in order to compute the relative distance and orientation to a

reference point. Sources for path integration include vestibular information, visual optical flow, and motor efference copies. For example, Mittelstaedt and Mittelstaedt [1980] showed that gerbils can perform angular integration and Etienne [1987] gives some evidence that they are able to perform path integration. For reviews on path integration see [McNaughton et al., 2006, Redish, 1999].

3.4.6 Models of Place Cells

Redish [1999] classifies place cell models as either purely allothetic local view models, or as including idiothetic cues. "Local view" is meant to include not only vision but any allothetic sensory cues, such as olfaction and audition, available from a certain spatial position and orientation. Such a model *"only depends on the local view to explain place cell firing"* [Redish, 1999]. Models of this class usually extract a number of features from sensory inputs in order to obtain a lower-dimensional representation that still carries information about spatial position in the environment but is invariant to everything else. Pure local view models do not comprise a path integration system and thus cannot fully explain oriospatial firing properties, e.g., in darkness. Pure path integration systems without external sensory input on the other hand accumulate errors, and hence a sensory coding mechanism is necessary to complement any such model.
Many place cell models have been proposed since the first finding of place cells. Reviews of these models can be found in [Redish, 1999, Andersen et al., 2007]. A detailed discussion of models related to this thesis is given in Section 4.4.1.

3.5 Head Direction Cells

Head direction cells were first found by Ranck Jr. [1985] (first full publication by Taube et al. [1990]) in the dorsal presubiculum, which is also known as the postsubiculum, of the rat. Later, head direction cells were also identified in many other brain areas including retrosplenial cortex, anterior thalamic nucleus, lateral dorsal thalamic nucleus, lateral mammillary nucleus, dorsal tegmental nucleus, striatum, and even some in the CA1 region of hippocampus proper [Taube and Bassett, 2003]. Firing of these cells is modulated by the animal's head direction relative to the environment in the horizontal plane (i.e., yaw) but mostly unaffected by pitch, roll, and location of the animal. The firing of a model head direction cell is illustrated in Figure 3.1 on page 27. The tuning curves of three representative head direction cells are plotted in Figure 3.3. Their function can be depicted as a compass, although in the rat they are not sensitive to geomagnetic field [Sharp et al., 2001] but rather to visual, tactile, and olfactory cues [Goodridge et al., 1998] and to vestibular input [Brown et al., 2002]. Head direction cells typically

have tuning curves with a single peak of activity but cases with two peaks of activity have been reported as well [Baird et al., 2001]. The orientation tuning curve normally is of Gaussian or triangular shape and shows no firing rate adaptation. The tuning widths lie in an interval of roughly 60° to 150° depending on brain area [Taube and Bassett, 2003, Sharp et al., 2001].

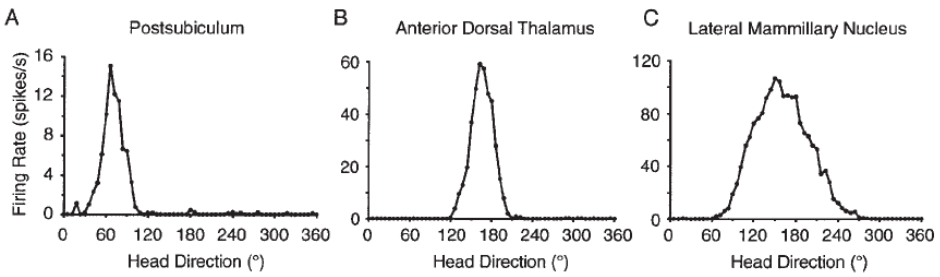

Figure 3.3: Tuning curves of three representative head direction cells. Data for these cells was recorded over 8 minutes sessions while the rat foraged for randomly placed food pellets about the floor of a cylindrical enclosure. Note that each cell has a different peak firing rate and preferred firing direction. The cell in C has a larger directional firing range compared to the other two cells. Cells were recorded in (A) postsubiculum, (B) anterior dorsal thalamus and (C) lateral mammillary nucleus. Figure reproduced from [Taube and Bassett, 2003] with permission from the publisher.

When the animal is confronted with a changed environment, the head direction cell system adapts to new visual stimuli after few minutes. Only 3 minutes after exposure to a new prominent visual cue, roughly half of the head direction cells lock to the position of the cue card in cue card rotation session and almost all cells do so after 8 minutes of exposure [Goodridge et al., 1998, Zugaro et al., 2003].

When an animal moves from one environment to another, head direction cells typically keep their preferred direction with respect to absolute direction. Obviously, this also means that in this case head direction cells keep their relative preferred directions, which is in contrast to the behavior of place cells, whose firing patterns are orthogonal in new environments. If, however, the rat is disoriented during the traversal, the cells often shift to new seemingly random preferred orientations in the new environment, but shift back to the "old reference frame" when reentering the first environment [Goodridge et al., 1998, Golob and Taube, 1997]. Head direction cells also change their preferred firing directions when the animal is passively transported to a new environment [Stackman et al., 2003]. When rats are disoriented during transport to another room, the preferred directions of

head direction cells can rotate coherently by a random angle. The head direction cells return to their previous preferred directions on reentry (or lights-on) of a known environment extremely quickly after approximately 100 ms. Some complex and possibly multimodal processing of sensory data has to be finished before such a "reset" can occur in this case, and this latency is much shorter than the approximately 150 ms required in humans or monkeys for object recognition tasks (i.e., the latency from stimulus onset to activity in the inferotemporal cortex, [Thorpe et al., 1996]). This experiment demonstrates that both angular integration and sensory cues influence head direction cell firing.

Head direction cell firing is unaffected by hippocampal lesions, even in novel environments [Golob and Taube, 1997]. When prominent cues are rotated coherently, head direction cells typically follow these cues. If multiple cues are incoherently changed, head direction cells almost always follow the more distal cues coherently as an ensemble [Yoganarasimha et al., 2006]. This fact is a strong indication for intensive coupling between head direction cells (see next section) and is in contrast to place cells (especially in CA1), which often split into subpopulations under similar conditions. The fact that head direction cells tend to bind to distal rather than local cues in case of cue conflict might be explained by the observation that distal cues serve as better landmarks because they roughly indicate a global direction (exactly if infinitely distant). When stroboscopic light perturbs the dynamic visual signals, this effect vanishes [Zugaro et al., 2004][14]. Although stroboscopic lighting will likely cause many sensory and behavioral changes, this finding supports the idea that spatial and temporal continuity of sensory information is necessary for head direction cell learning. This is in accordance with our model based on the temporal stability hypothesis in Chapter 4.
Head direction cells keep a constant firing rate when the animal ascends or descends wire-mesh walls in its cage vertically, but firing is disrupted when the animal moves invertedly (i.e., overhead) [Calton and Taube, 2005]. Stackman and Taube [1998] found head direction cells next to other cells with oriospatial correlates in the lateral mammillary nuclei. The authors describe "head pitch cells" that code for head pitch independent of head orientation and "angular velocity cells" that discharge as a function of angular head velocity in the horizontal plane.
The relation of the findings in this section with the model proposed in Chapter 4 is discussed in Section 4.4. Further discussions of head direction cell properties can be found for example in Wiener and Taube [2005] and Andersen et al. [2007].

[14] No such recordings seem to be published for any other oriospatial cell type.

3.5.1 Head Direction Cell Models

The most popular model class of head direction cells is a ring attractor network [reviewed for example in Touretzky, 2005]. This model class focuses on the angular integration and memory aspects of head direction cells. Head direction cells are modeled as a dynamical system consisting of a ring of mutually excitatory units. Under certain assumptions about weight distributions and nonlinear input-outputs functions of the units, the attractor of the system is one-dimensional and circular, meaning that any activity state on this one-dimensional manifold coding for head direction is stable even in the absence of sensory cues. Such a stable state is a roughly Gaussian-shaped "bump of activity" of adjacent units, which can be similar to the tuning curves of real head direction cells. In the absence of sensory cues this bump can slowly drift around the ring. Sensory cues enter the model in two ways: firstly, provided by visual feature detector cells and secondly, by "turn-modulated head direction cells" sensitive to angular velocity measured by sensors in the inner ear. The first cell class can reset the bump of activity if strong evidence for a certain orientation in space is provided by sensory cues (e.g., the presence of a cue card in a certain direction). How this information can be extracted from complex environments is not part of this model. However, given a set of feature detectors, simple Hebbian learning can be used to bind a feature detector for a specific direction to a head direction unit. This can, however, only work for highly distal cues, which is in accordance with the finding discussed above that the head direction cell system tends to prefer distal cues. The second class of sensory inputs is presumed to stem from two populations of neurons driven by head acceleration signals from the vestibular system. Each population forms an additional ring structure with similar bumps of activity as the head direction units and projects to the corresponding head direction unit. But during clockwise turns, one ring's bump is more active and slightly shifted in clockwise direction causing stronger inputs on the clockwise flank of the bump of activity in the head direction ring, which then causes it to move in that direction. The other ring of units is more active and shifted to the opposite direction for anticlockwise turns, causing a shift of the head direction cell's activity bump to the anticlockwise direction when the animal turns anticlockwise. Models of this class can explain how the head direction system keeps stable firing patterns in the absence of external cues and why a drift of preferred directions occurs after few minutes in darkness [Goodridge et al., 1998]. The ring attractor model can work instantaneously in new environments, in agreement with the findings discussed above. How the "visual feature detectors" should work, however, stays vague and cells of this type have not yet been found in the rodent brain [Touretzky, 2005]. Furthermore, in absence of highly distal cues (as in the high-walled cylinder) the simple feature binding approach would not result in a representation of global head

direction but instead in one like that of spatial view cells (see Section 3.7 and Figure 3.1). An alternative solution for the generation of such feature detectors is provided by the model proposed in Chapter 4.

3.6 Grid Cells

Grid cells are the most recently found major type of oriospatial cells in the hippocampal formation. They were recorded in layers 2/3 of the medial entorhinal cortex, which projects to the hippocampus proper. Grid cells were first described by Hafting et al. [2005]. Earlier recordings in the same region failed to unveil their spatial firing structure [e.g., Fyhn et al., 2004] since the standard recording environment of no more than one meter side length is too small to capture the spatial grid structure of the firing pattern that gives rise to the name of this cell type (see Section 3.2). Each grid cell fires in a remarkably exact hexagonal grid pattern with an average spatial distance between nearest maxima between 39 and 73 cm, depending on the recording site. Figure 3.1 on page 27 illustrates an idealized grid cell and Figure 3.4 depicts the firing behavior and autocorrelogram of a real grid cell. The spatial frequencies increase gradually and monotonically from ventral to dorsal in the medial entorhinal cortex (i.e., the spacing increases from dorsal to ventral EC). Grid cells recorded with the same tetrode have similar spatial frequencies and orientations but unrelated spatial phases, which seem to be homogeneously distributed [McNaughton et al., 2006, Hafting et al., 2005]. The grid frequencies seem to be independent of the size of the recording chamber. Recent reviews of grid cell properties are given in [McNaughton et al., 2006, Moser and Moser, 2007].

Unlike place cells whose firing fields typically change dramatically when sensory cues change (e.g., a "complete remapping" when a round recording chamber is exchanged with a square one), cells in the dorsal EC, which likely were grid cells, were mostly unaffected by such changes [Quirk et al., 1992]. Similar to place and head direction cells, firing patterns in grid cells remain stable in absence of sensory cues, suggesting the presence of path integration. More evidence for a strong influence of path integration on grid cell firing comes from the fact that the regular grid pattern seems to establish instantaneously after entry in a new room [Hafting et al., 2005]. But sensory cues also influence grid cell firing, indicated by the fact that their spatial phase in a given environment is consistent over multiple trials [Hafting et al., 2005]. Furthermore, Barry et al. [2007] have shown that grid cell firing is strongly influenced by visual cues and past experience. In their experiments, rats were placed in boxes with variable side lengths similar to the place field experiments by Muller and Kubie [1987] (see Section 3.2). After a rat had

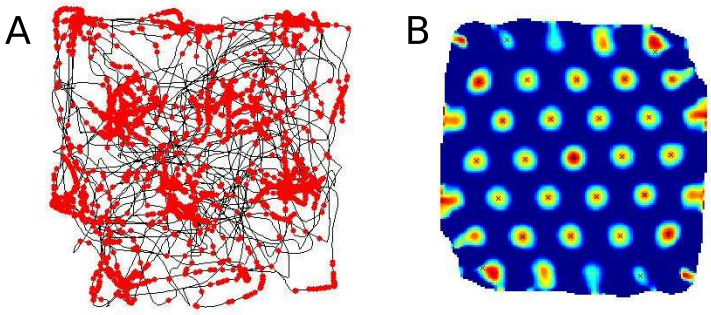

Figure 3.4: Firing map and spatial autocorrelogram of a grid cell. **A:** Gray lines indicate the rat's trajectory, red dots indicate positions where spikes occurred. **B:** Autocorrelogram of the same cell as in plot A. This plot illustrates the high regularity of the spatial firing structure of a grid cell. Source: T. Hafting, reprinted with permission from [Wikipedia, 2007].

become accustomed to a specific box geometry, the aspect ratio of the box geometry was changed, which lead to a significant deformation of the grid cell firing pattern. During repeated exposure to the new geometry, the grid slowly returned to the undeformed baseline condition. This experiment shows the joint influence of sensory cues and path integration on grid cell firing.

Similarly to place cells, some grid cells also correlate with other spatial variables apart from position only. In contrast to place cells, firing rates of few cells from layer 2 and many cells in layer 3 are modulated by head direction or running speed [Sargolini et al., 2006]. In summary, cells in medial entorhinal cortex may form a continuum between "pure spatial grid cells" and "pure head direction cells".

Because grid cells were discovered recently , many questions remain open. Does the grid-like firing pattern extend over larger and more complex areas? How regular does it remain under these conditions? Grid cells impose a specific metric on the environment that has not yet been experimentally investigated. Does a path on a tilted track cause the same firing pattern as in the plain? Do grids cells fire when the animal swims?

3.6.1 Grid Cell Models

The relatively young area of grid cell modeling has already inspired a number of models that will likely grow rapidly in the future. Most models focus on the emergence and advantages of distributed grid-like spatial codes performing path integration based on velocity signals [e.g., Burak et al., 2006, Fuhs and Touretzky, 2006, Guanella and Verschure, 2006] or how the grid

output might be used [e.g., Franzius et al., 2007b, Rolls et al., 2006, Solstad et al., 2006]. However, most models refrain from postulating mechanisms of sensory interaction with grid cells other than that some highly abstract signal should be used to bind grid cell firing to a specific phase in a given environment. O'Keefe and Burgess [2005] speculate that place cells, which could be anchored to a specific stimulus set perceived at the place field, might provide feedback to a specific grid cell and thus determine the firing phase of the grid cell. However, this postulate just shifts the complexity to the invariant stimulus recognition of place cells. The model proposed in Chapter 4 provides an alternative mechanism to generate a grid-like representation driven by sensory inputs.

3.7 Spatial View Cells

Spatial view cells in primates are correlated with the animal's position in space but show very different firing properties than place fields or head direction cells. These cells are neither position-invariant nor orientation-invariant but fire when the animal looks at a certain part of the environment (i.e., in eye-centered coordinates), irrespective of the animal's position in the room, resembling head direction cells for the case of an infinitely distant view. Figure 3.1 illustrates the spatial firing of a model spatial view cell. In this figure, a spatial view cell fires if the monkey is at any position in the room, as long as it fixates the view point marked by '×'. These cells are not simply triggered by the view of a specific object, which was tested by putting objects from the position of the '×' in front of the monkey without triggering the cell to fire [Rolls et al., 1997a]. Actual recordings from a macaque spatial view cell are depicted in Figure 3.5. Spatial view cells have not been found in rodents, whereas place and head direction cells occur in both primates and rodent. Ekstrom et al. [2003] also report spatial view cells (next to place and object cells) in the human hippocampus.

Experiments with primates are often different than rodent experiments: the primate is typically fixated in a static chair and watches stimuli on computer screens around it. In some more recent experiments the animal is put be in a mobile chair that the animal can control by itself (see Section 3.4 above). Are spatial view cells fundamentally different from place cells? Possibly spatial view cells encode the primate's notion of "space out there" [Rolls et al., 1997a] and thus the ability to identify places without physically being there, which would be a significant development from the rodent place field system that seems to code only for the animal's own momentary position. Alternatively, the properties of spatial view cells might be due to the changed

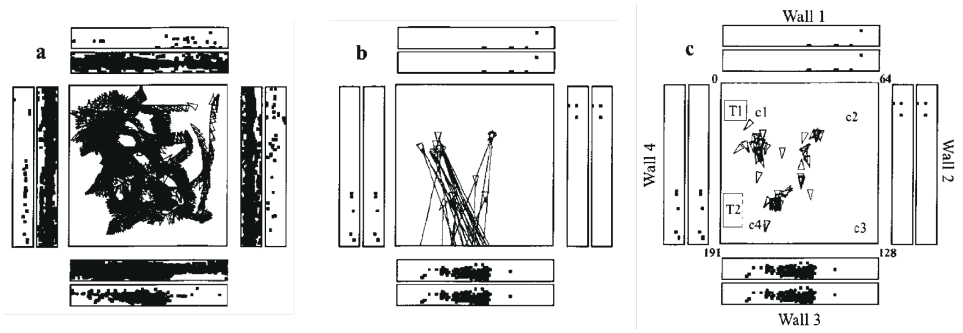

Figure 3.5: Examples of the firing behavior of a macaque hippocampal spatial view cell. The monkey was free to walk on all fours in a cart in the central 2.7 by 2.7 m of a room during three recording sessions of more than five minutes each. **a:** The firing of the cell is indicated by the spots in the outer set of four rectangles, each of which represents one of the walls of the room. The positions on the walls fixated during the recording sessions are indicated by points in the inner set of four rectangles. The central square is a plan view of the room, with a triangle printed every 250 ms to indicate the position of the cart, thus showing that many different places were visited during the recording sessions. **b**: A similar representation of the same three recording sessions as in A, but modified to indicate some of the range of cart positions and horizontal gaze directions when the cell fired. Sufficiently few cart/eye gaze direction icons so that they can be distinguished were selected by plotting only every 10th icon when the cell fired > 12 spikes/s. A spot was placed in the rectangles whenever the cell fired at > 12 spikes/s. **c**: A similar representation of the same three recording sessions as in B, but modified to indicate more fully the range of cart positions when the cell fired. Sufficiently few cart icons so that they can be distinguished were selected by plotting every cart icon when the cell fired > 12 spikes/s (12 spikes/s was selected as it was half the peak firing rate of the cell, and thus helps to reveal the conditions when the cell was strongly activated). Figure and caption reproduced from [Georges-François et al., 1999], with permission.

movement statistics of the animals in the experiments (i.e., fixation on a chair) and only an artifact of the artificial experimental setting. The model introduced in Chapter 4 could explain the different properties of place cells and spatial view cells based on such a difference in movement statistics. This question can only be solved by further experiments with freely behaving primates.

Similar to place cells that keep stable firing patterns in darkness, spatial view cells are not purely sensory driven as some cells exhibit sustained firing even when the fixated spot is hidden behind a curtain [Rolls, 1999]. This feature

indicates that some kind of memory-based spatial integration system, similar to path integration for place cells, is at work for spatial view cells.

3.7.1 Spatial View Cell Models

The model by de Araujo et al. [2001] suggests that the width of the field of view (FOV) is important for the distinction between spatial view cells and place cells. With a large FOV (as for rats) the animal can see most landmarks from all orientations while an animal with a small FOV (like many monkeys) can only see a subset of all landmarks at each point in time. In this model, cells are tuned to a specific distance of a specific set of (pointlike) abstract stimuli. A cell fires when at least three of its associated stimuli are in the correct distance within the agent's field of view. As a result, "rat" cells with a wide FOV become mostly orientation invariant, whereas "monkey" cells with a small FOV behave like spatial view cells. This model does, however, not account for the behavior-dependent differences in orientation specificity as discussed in Section 3.4 above. This article seems to be the only published model for spatial view cells.

3.8 Interactions Between Different Oriospatial Cells

The previous sections covered each oriospatial cell type individually. However, it is reasonable to assume that all these cells are part of a more general navigational framework with mutual interactions. This section reviews results from studies based on lesions and electrophysiology. However, much less data is available on interactions of different oriospatial cell types, as it often requires experimentally more demanding simultaneous recordings from different areas.

3.8.1 Interaction Between Place Cells and Head Direction Cells

The original formulation of the cognitive map theory stated that place and head direction cells together form a representation of positions, distances and directions between places [O'Keefe and Nadel, 1978]. A later extension of the theory in 1991 [reviewed in O'Keefe, 2007] suggested that head direction cells could give the necessary directional information. A necessary condition for such a model is that the place and head direction cell system stay in register after environmental changes like cue rotations. For coherent cue rotations, both place cells and head direction cells typically rotate in unison with the cue. The only simultaneous recording of place and head direction cells by Knierim et al. [1995] found both systems stayed in register even if sensory cues exerted no control over the cells, i.e., the seemingly random rotation occurred in unison.

Calton et al. [2003] found that lesions in postsubiculum (an area known to contain many head direction cells) reduced the orientation-invariance and spatial coherence of hippocampal place cells. Furthermore, the stimulus-dependence of the place cells was reduced: in unlesioned animals a single polarizing cue card controls the positions of place fields, whereas in the lesioned animals place fields shifted unpredictably. The authors conclude that input from head direction cells might be necessary for a fully functional place cell system. In the reversed scenario with a lesioned hippocampus no such effect on head direction cell firing was found [Golob and Taube, 1997] and thus head direction firing seems largely independent of a functional hippocampus. Both combined lesions in hippocampus and overlying neocortex or lesions only in the overlying neocortex have no large effect on head direction cell firing in a given known environment. When moving from one known environment to a new one, head direction cells in unlesioned animals maintain their preferred direction, an effect that is likely due to an angular path integration system. Head direction cells in the lesioned animals, however, did not maintain their preferred direction when traversing into the new environment [Golob and Taube, 1999], although cue card rotations still exerted control over the preferred directions in the new environment. Such lesions might thus be a key for dissociating stimulus-driven and path-integration-driven inputs for head direction cells.

3.8.2 Grid Cells and Place Cells

Lesions in the entorhinal cortex (EC) reduce the number of cells with spatial correlates in hippocampus and the robustness of the remaining firing fields and they strongly impair control of visual cues over place cell firing after cue rotations [Miller and Best, 1980]. Newer results by Steffenach et al. [2005] emphasize the necessity of the dorsolateral band of the EC for spatial learning in new environments. Combined with the fact that the EC provides major inputs to the hippocampus, it is highly likely that grid cells are critical for the formation of hippocampal place cells. However, as recall of previously established place fields remains stable after lesions in the dorsocaudal region of the EC, input from grid cells might not be necessary after place fields are established.

3.8.3 Place Cells and Visual Cortex

Save et al. [1998] report normal place cell firing in rats that were blinded early in their lives. In contrast to rats with damaged visual cortex, navigation performance in blind rats is not profoundly impaired [results reviewed in Poucet et al., 2003]. Poucet et al. [2003] therefore conjecture that the visual cortex might play an important role in spatial processing independently of actual visual input.

Chapter 4

A Model for Hippocampal Spatial Codes

This chapter introduces a model for the self-organized formation of place cells, head-direction cells, and spatial-view cells in the hippocampal formation based on unsupervised learning on quasi-natural visual stimuli. The model comprises a hierarchy of Slow Feature Analysis (SFA) nodes, which were recently shown to reproduce many properties of complex cells in the early visual system [Berkes and Wiskott, 2005]. The system extracts a distributed grid-like representation of position and orientation, which is transcoded into a localized place field, head direction, or view representation by sparse coding. The type of cells that develops depends solely on the relevant input statistics, i.e., the movement pattern of the simulated animal. Most of this chapter is based on the publication [Franzius et al., 2007a]. The numerical simulations are complemented by a mathematical analysis by Henning Sprekeler that allows us to accurately predict the output of the top SFA layer.

4.1 Experimental Methods

The outcome of an unsupervised learning rule, such as Slow Feature Analysis, is crucially determined by the statistics of the training data. As we want to show that oriospatial cells can be learned from raw sensory stimuli, we approximate the retinal stimuli of a rat by video sequences generated in a virtual-reality environment. The input statistics of the training data are thus jointly determined by the structure of the virtual-reality environment and the movement pattern of the simulated rat. As this video data is very high-dimensional, nonlinear SFA in a single step is computationally unfeasible. To overcome this problem, the model is organized as a hierarchy of SFA nodes in analogy to the hierarchy of the brain's visual system (see Figure 4.1C).

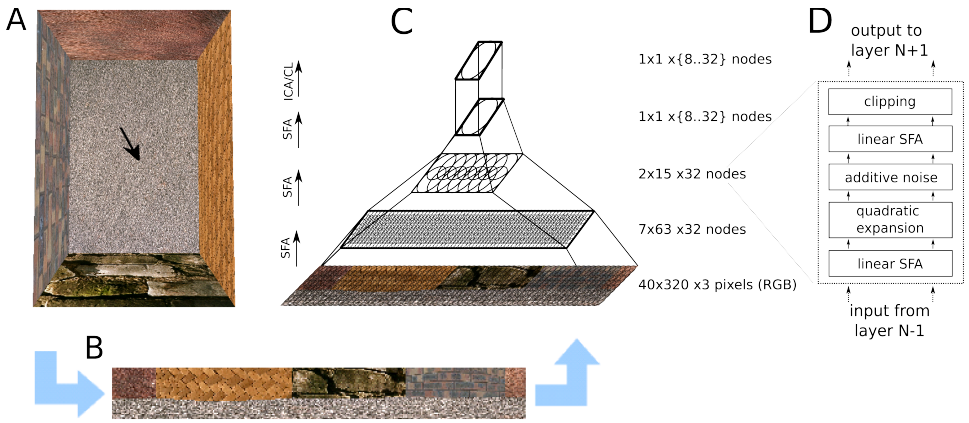

Figure 4.1: Architecture of the hierarchical model. At a given position and orientation of the virtual rat (arrow) in the naturally textured virtual-reality environment (A), input views are generated (B), and processed in a hierarchical network (C). The lower 3 layers perform the same sequence (D) of linear SFA (for dimensionality reduction), expansion, additive noise, linear SFA (for feature extraction), and clipping, the last layer performs sparse coding (either ICA or CL).

4.1.1 Simulated Environments

Many experimental place field data were recorded either in a linear track or in an open field apparatus. For our simulations we use a linear track of 10:1 side length and a rectangular open field of 3:2 side length. We have also simulated radial mazes (e.g., plus or 8-arm mazes) as a third apparatus type but they can be considered as a combination of an open field in the center with linear tracks extending from it and simulation results for this type will not be presented here.

The input data consists of pixel images generated by a virtual-reality system based on OpenGL with textures from the Vision Texture Database [Picard et al., 2002]. The virtual rat's horizontal field of view is 320° (see Figure 4.1A for a top view of the environment, and Figure 4.1B for a typical rat's view from this environment) and consistent with that of a real rat [Hughes, 1978]. The vertical field of view is reduced to 40° because outside this range usually only unstructured floor and ceiling are visible. An input picture has 40 by 320 color pixels (RGB, 1 pixel/°). The input dimensionality for the system is thus 38400, while the dimensionality of the interesting oriospatial parameter space is only three-dimensional (x- and y-position and orientation).

4.1.2 Movement Patterns of the Virtual Rat

As an approximation of a rat's trajectory during exploration in place field experiments we use Brownian motion on the three-dimensional parameter space of position and orientation (i.e., head direction). The virtual rat's position `pos(t)` at each time step `t` is updated by a weighted sum of the current velocity and Gaussian white noise `noise` with standard deviation `vr`. The momentum term `m` can assume values between zero (massless particle) and one (infinitely heavy particle), so that higher values of m lead to smoother trajectories and a more homogeneous sampling of the apparatus in limited time. When the virtual rat would cross the apparatus boundaries, the current velocity is halved and an alternative random velocity update is generated, until a new valid position is reached (see Table 4.1).

```
currentVelocity = pos(t) - pos(t-1);
repeat
   noise = GaussianWhiteNoise2d() * vr;
   pos(t+1) = pos(t) + m * currentVelocity + (1-m) * noise;
   if not isInsideApparatus(pos(t+1)):
      currentVelocity = currentVelocity / 2;
until isInsideApparatus(pos(t+1))
```

Table 4.1: Pseudocode for the computation of the translational movement of the virtual rat.

We call the standard deviation (normalized by room size L) of the noise term *translational speed* $v_{\mathbf{r}}$. In the *simple movement* paradigm the head direction is calculated analogously (but without checks for crossing of boundaries) and we call the standard deviation of the noise term (in units of 2π) for the head direction trajectory *rotational speed* v_ϕ. On long timescales and with finite room size this type of movement approximates homogeneous position and orientation probability densities, except at the apparatus boundaries where a high momentum term can increase the position probability. We call the ratio of rotational to translational speed v_ϕ/v_r the *relative rotational speed* v_{rel}. The actual choice of v_{rel} is based on the rat's behavior in different environments and behavioral tasks. In linear track experiments the rat's movement is essentially one-dimensional and the animal rarely turns on mid-track but instead mostly at the track ends. Accordingly, we use a large momentum term, so that the virtual rat often translates smoothly between track ends and rarely turns on mid-track. In the open field, on the other hand, full two-dimensional movement and rotation is possible, but the actual statistics depends on the behavioral task at hand. We mimic the common pellet-chasing experiment [Markus et al., 1995] by using isotropic two-dimensional translational speed and setting v_{rel} to a relatively high value.

In the simple movement paradigm head orientation and body movement are completely independent, so that head direction can be modeled with unrestricted Brownian motion. We also consider a *restricted head movement* paradigm, in which the head direction is enforced to be within ±90 degrees from the direction of body movement (see Table 4.2). This constraint implicitly restricts the range of possible relative speeds. While it is still possible to have arbitrarily high relative rotational speed by turning often or quickly, very low relative rotational speed cannot be achieved anymore in finite rooms. Typically, if the rat reaches a wall, it has to turn, resulting in a lower bound for the relative rotational speed v_{rel}. In order to generate input sequences with lower v_{rel} one needs to discard periods with dominant rotations from the input sequence. For a biological implementation of such a mechanism the rat's limbic system could access the vestibular rotational acceleration signal in order to downregulate the learning rate during quick turns. We will refer to this mechanism as *learning rate adaptation* (LRA). A third movement statistics can be generated if we assume that an animal

```
currentAngularVelocity = phi(t) - phi(t-1);
repeat
   noise = GaussianWhiteNoise1d() * vphi;
   phi(t+1) = phi(t) + m * currentAngularVelocity + (1-m) * noise;
until headDirIsWithin+/-90DegOfMovementDir(pos(t+1) - pos(t), phi(t+1))
```

Table 4.2: Pseudocode for the computation of the head direction of the virtual rat in the restricted head movement paradigm.

looks at objects or locations in the room for some time while moving around. During this period the animal fixates a specific location X in the room, i.e., it always turns its head into the direction of X, independently of its position. We implement X as a fixation point on the wall that moves in the following way: first, we generate an orientation ϕ using the algorithm in Table 4.2 and the same parameters as for the head-direction cell simulations. Second, the point X is defined as the point on the wall the rat would fixate if it were in the center of the room with head direction ϕ. We employ the identical translational movement mechanism as above, whereas the head direction is now completely determined by the animal's position and position of the viewpoint X. In this paradigm both position and orientation are dependent and vary rather quickly, while the position of X changes slowly. We call this movement pattern *spatial-view* paradigm and suggest that it is a more appropriate description of a primate's movement pattern than the previous two.

4.1.3 Model Architecture and Training

Our computational model consists of a converging hierarchy of layers of SFA nodes and a single final sparse coding step (see Figure 4.1C). Each SFA node finds the slowest output features from its input according to the SFA algorithm given in Section 2.2.2 and performs the following sequence of operations: linear SFA for dimensionality reduction, quadratic expansion with subsequent additive Gaussian white noise (with a variance of 0.05), another linear SFA step for slow-feature extraction, and clipping of extreme values at ± 4 (see Figure 4.1D). Effectively, a node implements a subset of full quadratic SFA. The clipping removes extreme values that can occur on test data very different from training data.

In the following, the part of the input image that influences a node's output will be denoted as its *receptive field*. On the lowest layer the receptive field of each node consists of an image patch of 10 by 10 pixels with 3 color dimensions each. The nodes form a regular (i.e., non-foveated) 7 by 63 grid with partially overlapping receptive fields that jointly cover the input image of 40 by 320 pixels. The second layer contains 2 by 15 nodes each receiving input from 3 by 8 layer 1 nodes with neighboring receptive fields, resembling a retinotopical layout. All layer 2 output converges onto a single node in layer 3, whose output we call *SFA-output*. Thus the hierarchical organization of the model captures two important aspects of cortical visual processing: increasing receptive field sizes and accumulating computational power at higher layers.

The network's SFA-output is subsequently fed into a final computational node that performs linear sparse coding, either by applying independent component analysis (we use CuBICA which is based on the diagonalization of third and fourth order cumulants [Blaschke and Wiskott, 2004]) or by performing competitive learning (CL). The top-layer output will be called *ICA-output*, or *CL-output*, respectively. ICA applied to non-localized grid-cell inputs finds sparser codes than CL, but the latter seems biologically more realistic. More details on different approaches for sparse coding of grid-cell input can be found in [Franzius et al., 2007b].

The layers are trained sequentially from bottom to top on different trajectories through one of the simulated environments. For computational efficiency we train only one node with stimuli from all node locations in its layer and replicate this node throughout the layer. This mechanism effectively implements a weight sharing constraint. However, the system performance does not critically depend on this mechanism. To the contrary, individually learned nodes *improve* the overall performance.

In analogy to a rat's brain, the lower two layers are trained only once and are kept fixed for all simulations presented here (like the visual system, which remains rather stable for adult animals). Only the top SFA and ICA layer are retrained for different movement statistics and environments.

For our simulations we use 100.000 time points for the training of each layer. Since training time of the entire model on a single PC is on the order of multiple days, the implementation is parallelized and training times thus reduced to hours. The simulated rat's views are generated from its configuration (position and orientation) with floating point precision and are not artificially discretized to a smaller configuration set.

The network is implemented in Python using the MDP toolbox [Berkes and Zito, 2005] and the code is available upon request.

4.1.4 Analysis Methods

The highly nonlinear functions learned by the hierarchical model can be characterized by their outputs on the three-dimensional configuration space of position and head direction. We will call two-dimensional sections of the output with constant (or averaged) head direction *spatial firing maps* and one-dimensional sections of the output with constant (or averaged) position *orientation tuning curves*. For the sparse coding results with ICA the otherwise arbitrary signs are chosen such that the largest absolute response is positive.

The sensitivity of a function f to spatial position r will be characterized by its mean positional variance η_r, which is the variance of $f(r, \phi)$ with respect to r averaged over all head directions ϕ: $\eta_r(f) = \langle \mathrm{var}_r(f(r, \phi)) \rangle_\phi$. Correspondingly, the sensitivity of a function f to head direction ϕ will be characterized by its directional variance η_ϕ averaged over all spatial positions r: $\eta_\phi(f) = \langle \mathrm{var}_\phi(f(r, \phi)) \rangle_r$. A perfect head-direction cell has no spatial structure and thus a vanishing η_r and positive η_ϕ, while a perfect place cell has positive η_r due to its spatial structure but no orientation dependence and thus a vanishing η_ϕ.

4.2 Theoretical Methods

Consider a rat in an environment that is kept unchanged for the duration of the experiment. The visual input the rat perceives during the experiment is the input signal for the learning task stated above. This section addresses the following question: Can we predict the functions learnt in such an experiment and, in particular, will they encode the rat's position in a structured way?

As the rat's environment remains unchanged for the duration of the experiment, the visual input cannot cover the full range of natural images but only the relatively small subset that can be realized in our setup. Given the environment, the rat's visual input can at all times be uniquely characterized by the rat's position and its head direction. We combine these parameters in a single *configuration vector* $\mathbf{s}$ and denote the image the rat perceives when it is in a particular configuration $\mathbf{s}$ as $\mathbf{x}(\mathbf{s})$. We refer to the

manifold of possible configurations as *configuration space* V. Note, that V in general does not have the structure of a vector space.

In a sufficiently complex environment we cannot only infer the image from the configuration but also the configuration from the image, so that there is a one-to-one correspondence between the configurations and the images. If we are not interested in how the functions the system learns respond to images other than those possible in the experiment, we can think of them as functions of the configuration $\mathbf{s}$, since for any function $\tilde{g}(\mathbf{x})$ of the images, we can immediately define an equivalent function $g(\mathbf{s})$ of the configuration:

$$g(\mathbf{s}) := \tilde{g}(\mathbf{x}(\mathbf{s})). \tag{4.1}$$

This leads to a simplified version of our problem. Instead of using the images $\mathbf{x}(t)$ we use the configuration $\mathbf{s}(t)$ as an input signal for our learning task.

It is intuitively clear that functions that vary slowly with respect to the configuration $\mathbf{s}$ will create slowly varying output when applied to $\mathbf{s}(t)$ as an input signal, because $\mathbf{s}(t)$ is continuous in time.

For the scenarios with homogeneous velocities and homogeneous spatial (or angular) probabilities of the animal, the optimal solutions are sinusoidal oscillations. Figure 4.2 shows the four basic cases of one- and two-dimensional optimal solutions under cyclic and free boundary conditions. The full mathematical analysis of optimal solutions by H. Sprekeler is provided in [Franzius et al., 2007a]. An additional insight of this analysis is that the optimal functions show oscillations that are spatially compressed in regions where the rat moves with low velocities. This implies that the spatial resolution of the SFA solutions is higher in those regions. Consequently, the size of the place fields after sparse coding should be smaller in regions with small velocities, which might explain smaller place fields near arena boundaries [Muller et al., 1987, O'Keefe, 2007]. If we assume the animal moves faster parallel to a wall of the arena than perpendicular to it, our theory predicts elongated place fields along the walls that might be similar to the crescent-shaped fields reported in [Muller, 1996] for a circular arena.

4.3 Results

We have applied our theoretical framework and computer simulations to a number of environments and movement patterns that resemble typical place cell experiments. In Section 4.3.1, we show results for the open field, beginning with the mathematical analysis and simulation results for the simple movement paradigms with high and low relative speeds. Subsequently, the simulation results for the restricted head movement paradigm, including learning rate adaptation, and the spatial-view paradigm are shown. In Section 4.3.2 the results for the linear track with its two-dimensional configuration space are shown.

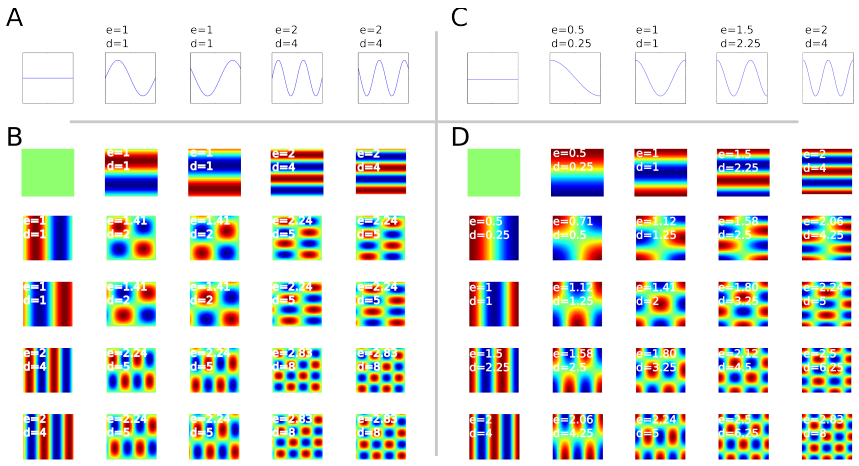

Figure 4.2: Illustration of optimal solutions for SFA for homogeneous velocities and probabilities of two features x_1 and x_2. Examples for the one- and two-dimensional cases, each for free and cyclic boundary conditions are shown. x_1 and x_2 are assumed to be uncorrelated and have identical Δ-values and a domain of $[0, 2\pi]$. Insets denote Δ-values (d) and η-values (e). **A:** One-dimensional signal with cyclic boundary conditions. Solutions are of the form $y_k(x) = \sin(kx)$ (for k even) and $y_k(x) = \cos(1 + (k-1)kx)$ (for k odd). **B:** Two-dimensional signal with cyclic boundary conditions. Solutions are products of those in panel A. **C:** One-dimensional signal with free boundary conditions. Solutions are of the form $y_k(x) = \cos(\pi kx/2)$. **D:** Two-dimensional signal with free boundary conditions. Solutions are products of those in panel C.

4.3.1 Open Field

One of the most common environments for place cell experiments is an open field apparatus of rectangular or circular shape. Here, the most typical experimental paradigm is to throw food pellets randomly into the apparatus at regular intervals leading to a random search behavior of the rat. For this case the rat's oriospatial configuration space comprises the full three dimensional manifold of position and orientation. In this section, we present results from experiments with simulated rat trajectories at either high or low relative rotational speeds leading to undirected place cells or position-invariant head-direction cell type results, respectively.

Theoretical Predictions for the Simple Movement Paradigm

In a rectangular open field the configuration space can be parametrized by the animals position, indicated by the coordinates x and y, and its head direction ϕ. The total configuration space is then given by $\mathbf{s} = (x, y, \phi) \in [0, L_x] \times [0, L_y] \times [0, 2\pi[$. L_x and L_y denote the size of the room in x- and y-direction, respectively. We choose the origin of the head direction ϕ such that $\phi = \pi/2$ corresponds to the rat looking to the North. The velocity vector is given by $\mathbf{v} = (v_x, v_y, \omega)$, where v_x, v_y denote the translation velocities and ω is the rotation velocity. For the typical pellet-throwing experiment we make the approximation that the velocities in the three different directions are decorrelated and that the rat's position and head direction are homogeneously distributed in configuration space. Moreover, in an open field there is no reason why the variance of the velocity should be different in x- and y-direction. Let

$$ {v_{\mathrm{rel}}}^2 = \frac{\langle (\frac{\omega}{2\pi})^2 \rangle}{\langle (\frac{v}{L_x})^2 \rangle} \tag{4.2} $$

denote the relative rotational speed, i.e., the ratio of the root mean square of rotational and translational velocity, if translational velocity is measured in units of the room size in x-direction per second and rotational velocity is measured in full circles per second.

We can now discuss two limit cases in terms of the relative velocity v_{rel}. Let us first consider the case where the rat moves at small velocities while making a lot of quick turns, i.e., $v_{\mathrm{rel}} \gg 1$. In this case, the smallest Δ-values can be reached by functions that do not depend on the angle ϕ at all. The slowest functions for this case are invariant with respect to head direction and lead to place cells, see below. The behavior of the solutions and the respective simulation results are depicted in Figure 4.3A and B.
In the other extreme, v_{rel} is much smaller than one, i.e., the rat runs relatively fast while making few or slow turns. The smallest Δ-values can then be reached by functions that completely ignore the position. These functions are invariant with respect to position while being selective to head

direction, a feature that is characteristic for head-direction cells. A comparison of these theoretically predicted functions with simulation results are shown in Figure 4.3D and E. The full mathematical derivation of these two cases is given in [Franzius et al., 2007a].

Figure 4.3: (see page 64) **Theoretical predictions and simulation results for the open field with the simple movement paradigm (independent translation and head direction), separately learned place cells and head-direction cells, and ICA for sparsification.** Each row within each panel shows the response of one unit as a function of position for different head directions (indicated by arrows), as well as the mean value averaged over all head directions (indicated by the superimposed arrows). Blue color denotes low, green intermediate, and red high activity. Panel C also shows orientation tuning curves at the position of a unit's maximal activity. Panels D-F also show orientation tuning curves averaged over all positions ± one standard deviation.
A: Theoretical prediction for the SFA layer with relatively quick rotational speed compared to translational speed ($v_{\text{rel}} = 32$). Solutions are ordered by slowness. All solutions are head direction invariant and have regular rectangular grid structures. **B:** Simulation results for the SFA layer for the same settings as in A, ordered by slowness. The results are similar to the theoretical predictions up to mirroring, sign, and mixing of almost equally slow solutions. All units are head direction invariant and code for spatial position but are not localized in space. **C:** Simulation results for the ICA layer for the same simulation as in B, ordered by sparseness (kurtosis). Firing patterns of all units are head direction invariant and localized in space, resembling hippocampal place cells. **D:** Theoretical prediction for the SFA layer for relatively slow rotational speed ($v_{\text{rel}} = 0.08$) compared to translational speed. Solutions are ordered by slowness. All solutions are position invariant and constitute a Fourier basis in head direction space. As the phases of these theoretical solutions are not uniquely determined, they were adjusted to match the simulation results in E. **E:** Simulation results for the SFA layer for the same settings as in D, ordered by slowness. The results are similar to the theoretical predictions. All units are position invariant and head direction specific but not localized in head direction space, i.e., all units except 1 and 2 have multiple peaks. **F:** Simulation results for the ICA layer for the same simulation as in E ordered by sparseness (kurtosis). Firing patterns of all units are position invariant and localized in head direction space resembling subicular head-direction cells.

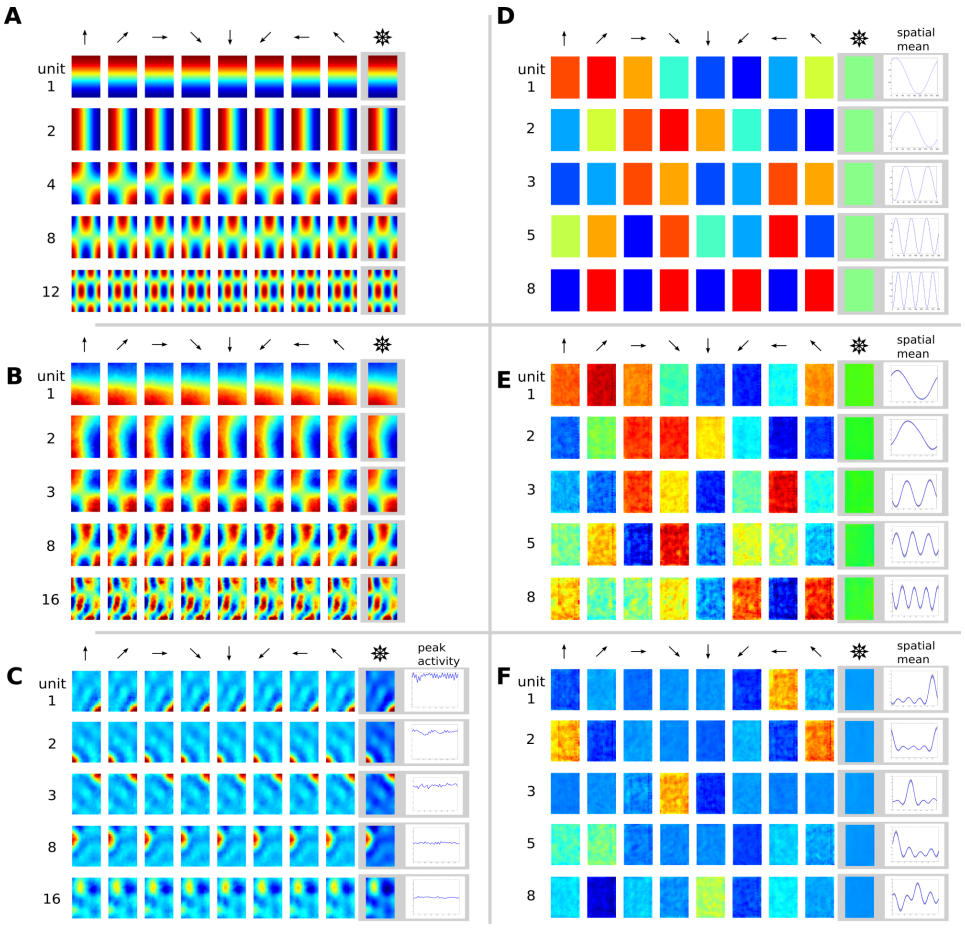

Figure 4.3: Caption on page 63.

Simulation Results for the Simple Movement Paradigm

It is clear that for high relative orientational speed v_{rel} the system output becomes slowest if it is invariant to head direction and only codes for spatial position. For low v_{rel} on the other hand invariance for position while coding for head orientation is the best solution to the optimization problem.

In Figure 4.3B the spatial firing maps of SFA output units from the simulation with high $v_{\text{rel}} = 32$ are shown. Here, all units are almost completely orientation-invariant and resemble the theoretical predictions from Figure 4.3A. The first unit has low activity when the simulated rat is in the South of the apparatus, is most active in the North, and shows a gradual increase in the shape of a half cosine wave in between. The unit is invariant to movements in East-West direction. The second unit behaves similarly, but its activity pattern is rotated by 90 degrees. The following units have more spatial oscillations and somewhat resemble grid cells, which are not localized.

Figure 4.3C shows ICA output units from the same simulation as in Figure 4.3B. All units are orientation-invariant, just as their input from the first 16 SFA units, but most have only a single peak of activity and each at a different position. The sparser units are more localized in space while less sparse units have larger firing fields or multiple peaks. These results closely resemble place cells from rodent's hippocampal areas CA1 and CA3.

In Figure 4.3E SFA output units from the simulation with low relative rotational speed $v_{\text{rel}} = 0.08$ are shown. In this case, all units are almost completely position-invariant but their response oscillates with the orientation of the rat. The first unit changes activity with the sine of orientation and the second unit is modulated like a cosine. Unit 3 has twice the frequency, unit 5 has a frequency of three, and unit 8 a frequency of four. Again, the simulation results reproduce the theoretical predictions shown in Figure 4.3D. Figure 4.3F shows ICA output units from the same simulation as in Figure 4.3E. All units are position-invariant like their inputs from the first 8 SFA units, but most have only a single peak of activity and each at a different orientation. The sparser units are more localized in orientation while later ones have broader tuning curves. These results closely resemble head-direction cells from rodent's subicular areas.

Simulation Results for the Restricted Head Movement Paradigm

In the previous section we used independent head direction and body movement and used different movement statistics for different cell types, such as fast rotational speed for place cells and slow rotational speed for head-direction cells. This allowed us to obtain nearly ideal simulation results that match closely the theoretical predictions, but it is unrealistic for two reasons. Firstly, in a real rat head-direction and movement direction are correlated. Secondly, in a real rat place cells and head-direction cells have to be learned simultaneously and thus with the same movement pattern.
In this section we introduce three changes for higher realism. Firstly, a more realistic movement pattern is used, where the rat's head is enforced to be within $\pm 90^\circ$ of the current body movement (see methods) and the relative rotational speed v_{rel} is set to an intermediate value of 0.6. Secondly, place cells and head-direction cells are learned on the same input statistics and learning rate adaptation (LRA) is used in the top SFA layer for the head-direction cell population (see methods). Thirdly, ICA for sparse coding in the last layer is replaced by competitive learning (CL). Simulation results are shown in Figure 4.4.

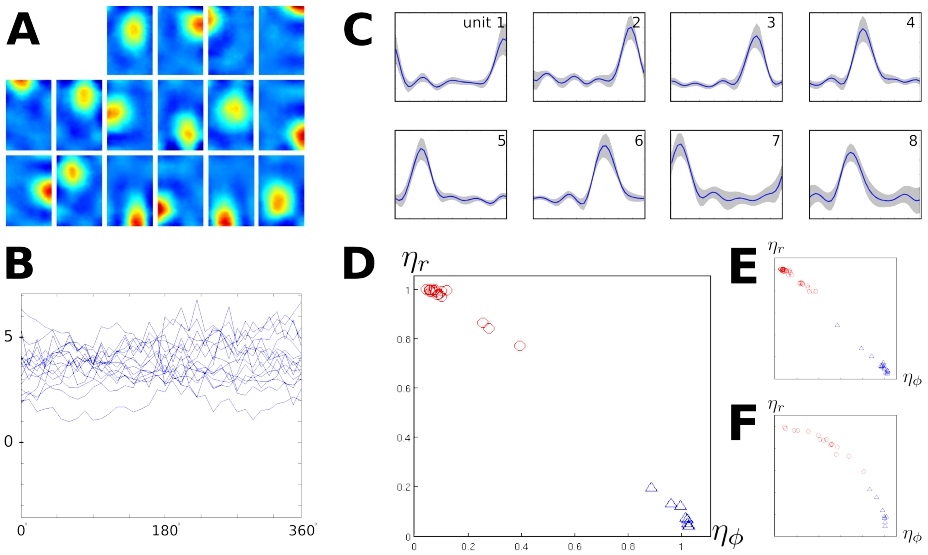

Figure 4.4: Caption on page 67.

As the relative rotational speed v_{rel} is smaller than in the previous section some SFA solutions (not shown) change with head direction: unit 16 of 32 is the first unit with noticeable head direction dependence here while none of the first 32 SFA solutions in the place cell simulation in the last section was head direction dependent. In Figure 4.4A the spatial firing maps for all units trained without LRA are shown averaged over all orientations. The corresponding orientation tuning curves (measured at the peak of the place field) are given in panel B. All units are localized in space and largely independent of orientation with activity centers distributed evenly in the room.

Figure 4.4C shows the simulation results with identical movement statistics but with LRA turned on in the top SFA layer, so that learning is down-regulated at time points with rapid head direction changes. Tuning curves of all units are shown together with the spatial standard deviation of activity, which is generally very small. All units are localized in head direction space and mostly position independent with approximately even spacing of directions of maximum activity. The LRA can eliminate the effect of head rotation only to some extent and thus SFA units 7 and 8 (not shown) show significant dependence on position while the slowest unit affected by position in the previous section was unit 15.

A scatterplot of the mean positional variance η_r versus mean orientational variance η_ϕ (see methods) of the units from A and C is shown in Figure 4.4D. Perfect head-direction cells would be located in the bottom right while perfect place cells would be located in the top left. Red circles denote the simulated place cells from panel A; the blue triangles denote the

Figure 4.4: (see page 66) **Simulation results for the open field with more realistic movement patterns and competitive learning (CL) for sparsification in the last layer.** The network was trained with a movement pattern of relatively high rotational speed. Two distinct populations of cells were trained, one as before, the other was trained with learning rate adaptation (LRA) in the top SFA layer, reducing the impact of periods with high rotational speed.
A: Simulation results for the top layer CL units without LRA. Each subplot shows the mean spatial firing rate of one output unit averaged over all orientations. The slowest 16 SFA outputs were used as an input for CL, and 16 CL units were trained. All units are localized in space, closely resembling hippocampal place cells. Blue color denotes low, green intermediate, and red high activity. **B:** Orientation tuning of the units shown in A. Firing patterns of all units are mostly head direction invariant. **C:** Simulation results for the top layer CL units with LRA in the top SFA layer. Each subplot shows the mean orientation tuning curve in blue and a gray area indicating ± one standard deviation. The slowest 8 SFA-outputs were used for CL, and 8 CL units were trained. Firing patterns of all units are mostly position invariant and localized in head direction space closely resembling subicular head-direction cells. **D:** Scatterplot of mean directional variance η_ϕ and mean positional variance η_r for the results shown in A (red circles) and C (blue triangles). Units from A cluster in an area with high positional variance η_r and low orientational variance η_ϕ, while units from C cluster in an area with low positional variance η_r and high orientational variance η_ϕ. **E:** Scatterplot of η_ϕ and η_r for the same simulation parameters as in A-D but with more CL output units. 32 units were trained without LRA (red circles) and 16 with LRA (blue triangles). The solutions lie in similar areas as in D. **F:** Scatterplot of η_ϕ and η_r for the same simulation parameters as in A-D, but with more SFA outputs used for CL. 32 SFA units were used without LRA (red circles) and 16 with LRA (blue triangles). The solutions show mixed dependence on position and head direction but are still clearly divided into a mostly head direction-invariant population (red) and a mostly position-invariant population (blue).

simulated head-direction cells from panel C. Both populations cluster near the positions of optimal solutions in the corners.

How does the number of inputs to the last layer (i.e., the number of SFA-outputs used) and the number of CL outputs influence the results? Panel E shows the same analysis for a simulation with identical settings except the number of CL-output units was doubled to 32 without LRA and 16 with LRA, respectively. Most units lie in a similar area as in D, but the clusters are denser, since the number of units has doubled. In panel F, the number of output units is again the same as in D, but the number of

SFA outputs for the last layer is doubled to 32 for the simulation without LRA and 16 for the simulation with LRA. The output units now get inputs from higher, i.e., quicker, SFA units which tend to depend on both position and orientation. As a result, the CL units span the entire spectrum of completely position invariant to complete orientation invariant solutions, with the more position-dependent solutions coming from the simulations without LRA, and the more head-direction dependent solutions coming from the LRA simulation. We have no conclusive explanation, though, why the shape of the data distribution seemingly changes from linear (panels D,E) to convex (panel F) with increasing numbers of SFA units. We conclude that the number of CL-output units mostly determines the density of place cells but not the qualitative behavior of the solutions while the number of SFA-outputs directly affects the invariance properties of the solutions.

Simulation Results for the Spatial-View Paradigm

The previous sections have shown that the same learning mechanism in the same environment, just with different movement statistics, results in either head-direction or place cell like representations. Although the last section introduced certain restrictions on the head direction, body position and head direction remained mostly independent.

In the following simulation, the virtual animal fixates a location X on a wall while it moves through the room. The position of X is subject to a random walk on the wall with the same statistics as the head direction in the simple movement paradigm with small v_{rel} (see methods). The animal's position is also changed with the same statistics as position in the simple movement paradigm, and the actual head direction is thus determined by the current position and currently fixated point X.

Note that the configuration space consisting of position and view point has the same structure as the one consisting of position and head direction for the simple movement paradigm. Accordingly, the theoretical predictions for the two scenarios are identical if head direction is "replaced" by the fixation point. In Figure 4.5C we plot the spatial activity pattern such that at each position the rat fixates a specific location marked by an '$\times$'. As expected, these plots are virtually identical to the head direction cell plots in Figure 4.3D-E in that activity is largely invariant to position. This can also be seen by the corresponding tuning curves that show small standard deviations (indicated by gray areas). However, while in Figure 4.3D-E the activities are modulated by head direction, activities in plot 4.5C depend on the position of view point. If we plot the same data with fixed head direction instead of fixed view point (plot 4.5A), the structure of the activity patterns is obscured. Units 3-5 in Figure 4.5A, for example, show clear diagonal stripes and correspondingly larger standard deviations in their tuning curves.

These SFA solutions jointly code for 'view space' but as before the SFA results are not localized. Figure 4.5B and D show the results of the ICA layer. The 'global direction' plot in B is as inadequate as in A while plot D clearly illustrates the behavior of these cells. Unit 2, for example, is active only when the rat looks at the bottom left corner of the rectangular room, independently of the animal's position. This cell type resembles spatial-view cells found in the primate hippocampal formation [e.g., Rolls et al., 2005].

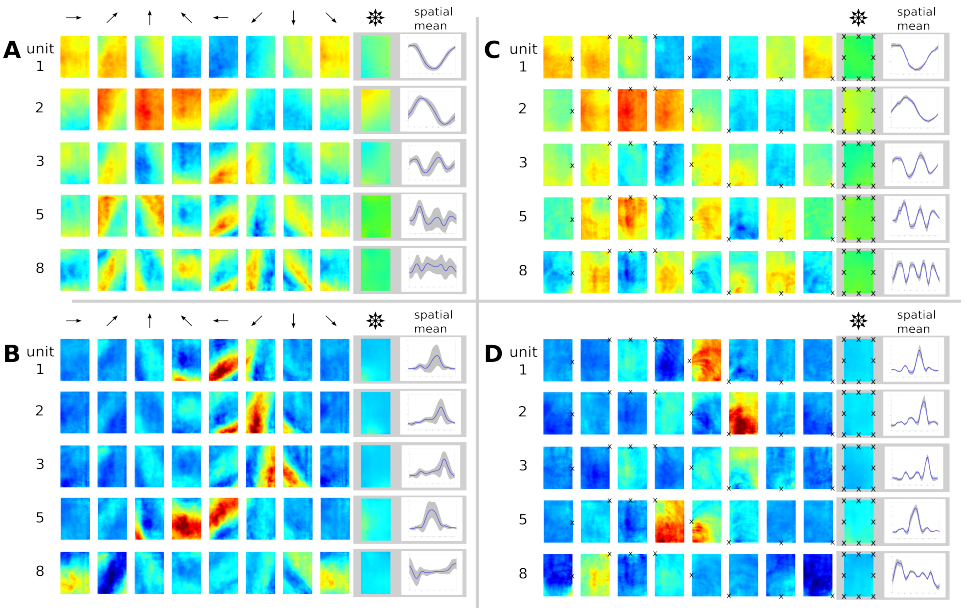

Figure 4.5: Simulation results for the open field with trajectories where spots on the wall were fixated. Blue color denotes low, green intermediate, and red high activity. **A:** Spatial firing map of five representative SFA output units for different 'global head directions' (indicated by arrows) and averages over orientation and space. No unit shows spatial or orientation invariance when plotting position and 'global head direction' as in previous figures. **B:** ICA results plotted with 'global head direction'. **C:** Same results as in A but plotted with 'local head direction' (at each position oriented towards fixation point '×'). **D:** Same results as in B but using the plot method from C. All units code for a specific view closely resembling primate spatial-view cells.

4.3.2 Linear Track

In a linear track the rat's movement is essentially restricted to two degrees of freedom, a spatial and an orientational one. In experimental measurements the orientational dimension is often collapsed into a binary variable indicating only the direction of movement. In the linear track these two

dimensions are thus experimentally much easier to sample smoothly than the full three dimensional parameter space of the open field.

Theoretical Predictions for the Linear Track

In principle the configuration space for the linear track is the same as for the open field, only with a small side length L_x in one direction. For small L_x the solutions that are not constant in the x-direction, i.e., the solutions with $l \neq 0$, have large Δ-values and thus vary quickly. Therefore slow functions will be independent of x and we will neglect this dimension and restrict the configuration space to position in y-direction and head direction ϕ.

Another difference between the simulation setup for the open field and the linear track lies in the movement statistics of the rat. Due to the momentum of the Brownian motion the rat rarely turns on mid-track. In combination with the coupling between head direction and body motion this implies that given the sign of the velocity in y-direction the head direction is restricted to angles between either 0 and π (positive velocity in y-direction, North) or between π and 2π (negative velocity in y-direction, South). If, in addition, the rat makes a lot of quick head rotations, the resulting functions can only be slowly varying if they are invariant with respect to head direction within these ranges. This leaves us with a reduced configuration space that contains the position y and a binary value $d \in \{\text{North, South}\}$ that determines whether $0 \leq \phi < \pi$ or $\pi \leq \phi < 2\pi$.

We assume that the rat only switches between North and South at the ends of the track. Because discontinuities in the functions lead to large Δ-values, slow functions $g(y, d)$ should fulfill the continuity condition that $g(0, \text{North}) = g(0, \text{South})$ and $g(L_y, \text{North}) = g(L_y, \text{South})$. This means that the configuration space has the topology of a circle, where one half of the circle represents all positions with the rat facing North and the other half the positions with the rat facing South.

Note that there are always two functions with the same Δ-value. Theoretically, any linear combination of these functions has the same Δ-value and is thus also a possible solution. In the simulation, this degeneracy does not occur, because mid-track turns do occur occasionally, so those functions that are head-direction-dependent on mid-track (i.e., even j) will have higher Δ-values than theoretically predicted. This avoids mixed solutions and changes the order of the functions when ordered by slowness. Figure 4.6A shows seven of the theoretically predicted functions g_j, reordered such that they match the experimental results.The full mathematical derivation of this section can be found in [Franzius et al., 2007a].

Simulation Results for the Linear Track

For simulations in the linear track we use a restricted head movement paradigm similar to that of the open field experiment from Section 4.3.1. A similar relative speed is assumed ($v_{\text{rel}} = 26$) and sparse coding in the last layer is performed with ICA.

Figure 4.6B and C shows the simulation results for the linear track. The spatial firing maps of the seven slowest SFA outputs out of ten are shown in Figure 4.6B. Units 1–6 are mostly head direction invariant ($\eta_\phi \leq 0.1$), and code for spatial position in the form of sine waves with frequencies of $\frac{1}{2}$, 1, $1\frac{1}{2}$, 2, $2\frac{1}{2}$, and 3, as theoretically predicted. Units 7–10 (latter three not shown) code for position and orientation. At track ends, where most rotation occurs, all units are head-direction invariant and the spatial modulation is compressed due to slower mean translational speeds compared to mid-track (cf. Section 4.2). As expected, none of these units are localized in space or orientation.

The spatial firing maps of the first seven out of ten ICA outputs for different head directions are shown in Figure 4.6C. Units 1 and 6 are only active at the southern track end independently of head direction. Units 9 and 10 (not shown) are active on mid-track and mostly independent of head direction ($\eta_\phi \leq 0.1$).The other six units are localized in the joint position-head-direction space meaning that they fire only at specific positions on the track when the rat faces a specific direction. These results are similar to place cell recordings from rats in linear tracks where most cells only fire when the rat moves in one direction [Muller et al., 1994].

Changing the movement pattern to yield much higher or much lower mean relative rotational speeds can lead to very different results resembling those presented earlier for the open field, namely head-direction cells and head-direction invariant place cells.

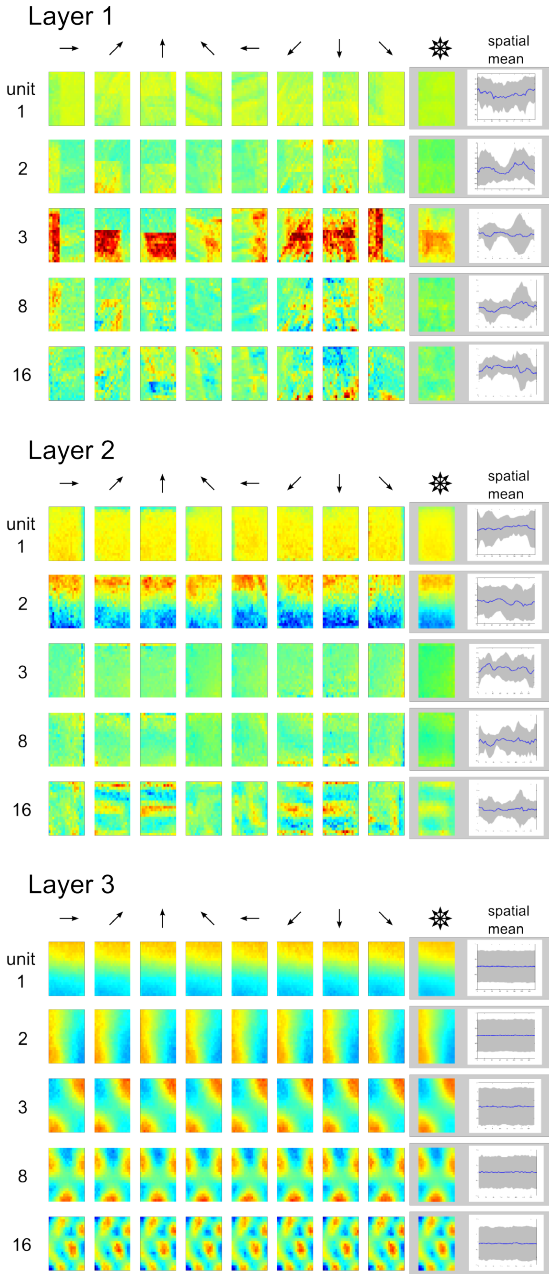

Figure 4.7: Caption on page 74

4.3.3 Model Parameters

Although most of the parameters in our model (i.e., all the weights in the SFA and ICA steps) are learned in an unsupervised manner, a number of parameters were chosen by hand. These parameters include the input picture size, receptive field sizes, receptive field positions and overlaps in all

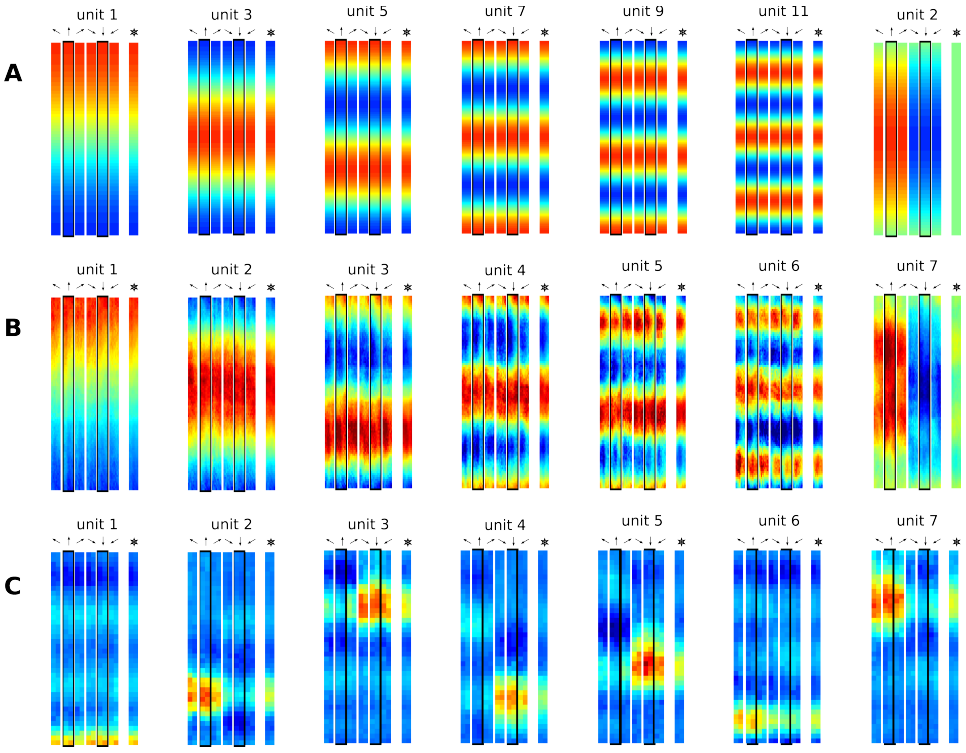

Figure 4.6: Theoretical predictions and simulation results for the linear track. Head directions are indicated by arrows, orientation averages are indicated by superimposed arrows, and principal directions (North, South) are emphasized with a dark border. Blue color denotes low, green intermediate, and red high activity. **A:** Theoretical predictions. **B:** Spatial firing maps of the first (i.e., slowest) seven out of ten SFA output units. Units 1–6 are mostly head direction invariant, whereas unit 7 responds differently to North and South views. Two out of the three remaining units are also head direction invariant. **C:** Spatial firing maps of the first (i.e., most kurtotic) seven out of ten ICA output units. All units are localized in space and most of them are only active for either North or South views closely resembling place fields recorded from rats in linear track experiments.

layers, the room shape and textures, the expansion function space, number of layers, choice of sparsification algorithm, movement pattern, field of view, and number of training steps. We cannot explore the entire parameter space here and show instead that the model performance is very robust with respect to most of these parameters. The fact that the presented simulation results are very similar to the analytical solutions also indicates that the results are generic and not a mere artifact of a specific parameter set. The most interesting parameters are discussed in the following:

Figure 4.7: (image on page 72) Simulation results for the open field with the simple movement paradigm (independent translation and head direction), with high v_{rel} from the three SFA layers. Each row within each panel shows the response of one unit as a function of position for different head directions (indicated by arrows), as well as the mean value averaged over all head directions (indicated by the superimposed arrows). Blue color denotes low, green intermediate, and red high activity. Orientation tuning curves show averages over all positions ± one standard deviation. The top panel shows activity maps of units from one central position in the first layer. All activities strongly depend on position (maximal $\eta_r = 0.99$, average $\eta_r = 0.95$) and orientation (maximal $\eta_\phi = 0.97$, average $\eta_\phi = 0.87$). The middle panel shows activity maps of units from one central position in the second layer. All activities depend on position (maximal $\eta_r = 1.0$, average $\eta_r = 0.99$) and orientation (maximal $\eta_\phi = 0.91$, average $\eta_\phi = 0.58$). The lowest panel shows activity maps from units in the third (and highest) SFA layer, which are similar to Figure 4.3B. These results are close to the theoretically predicted optimal solutions and specifically are mostly independent of head direction (maximal $\eta_\phi < 0.04$, average $\eta_\phi = 0.02$) but depend on position (maximal $\eta_r = 1.00$, average $\eta_r = 1.00$).

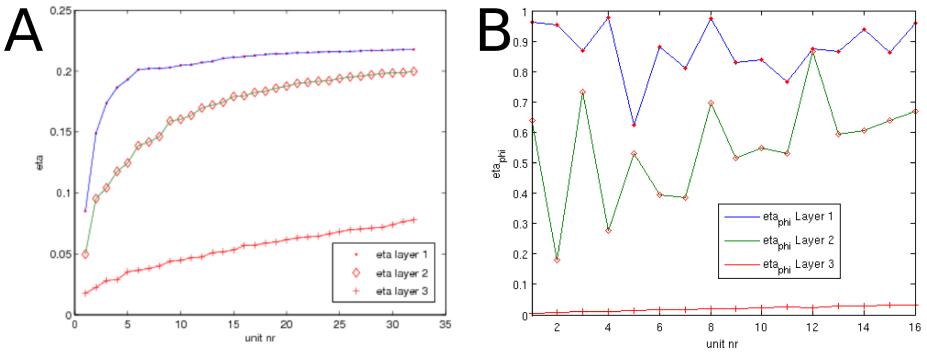

Figure 4.8: η-values and orientation dependencies (η_ϕ-values) per network layer. A: Eta value distributions of a central unit from three SFA layers on training data corresponding to Figure 4.7. Higher layers extract slower features. **B:** η_ϕ-values from the same units as in A. Units in the first two layers show high degrees of orientation sensitivity, whereas units in layer 3 are mostly invariant to orientation. Furthermore, orientation dependency increases slowly and monotonically with increasing Δ-value in layer 3. Additional top layers show similar behavior as layer 3 (data not shown).

Image Resolution: We use high-resolution input pictures of 40 by 320 RGB pixels showing the capability of the model to handle high-dimensional

sensory data. However, it could be argued that the rat's vision is rather blurred and has little color sensitivity. We find that smaller and/or grayscale input pictures yield similar results, which degrade only below a dimensionality of a few hundred input pixels.

Field of view: The model's field of view (FOV) has been modeled to represent the 320° of a rat's FOV. Smaller FOVs down to 60° still reproduce our results and especially rotation invariance is not an effect of a large FOV. However, the views have to contain enough visual information in order to fulfill the one-to-one correspondence between stimulus and oriospatial configuration.

Receptive Fields: The receptive fields are restricted to about 100 input dimensions (before quadratic expansion) due to computational limitations. Larger receptive fields tend to yield better solutions, since the available total function space increases. Position and overlap of receptive fields have been varied to some extent but have no noticeable impact on the result unless too many of the inputs are discarded.

Room shape: The room shape has a strong impact on the SFA solutions, which can be predicted analytically. We show here only results for convex rooms, but experiments with radial mazes and multiple rooms have been performed and these results are similar to experimental data, too. Figure 4.9 shows additional simulation results for a circular room. Choice of specific textures was irrelevant for the model's performance except when multiple walls are textured with similar or identical textures, which leads to degraded results due to visual ambiguities. For small FOV values and symmetrical environments the model's representations become symmetrical as well.

Nonlinear expansion: The expansion function space was chosen as all monomials up to degree 2, but alternative function spaces like linear random mixtures passed through sigmoidals with different offsets were successful, too. However, the size of the function space is limited by computational constraints and monomials have proven to be particularly efficient. Even a linear function space is sufficient to generate a subset of the theoretically predicted results in some cases. The head-direction cell simulations reproduce 7 out of 8 optimal SFA solutions in the linear case and with a 320° FOV. In a linear place cell simulation only every second optimal SFA solution was found and most of the ICA representations had two or more separate peaks. Simulations with a linear function space yield the theoretically predicted results only for a large FOV.

Number of layers: The number of layers is determined by receptive field sizes and overlaps. An increased number of layers also increases the

function space and can thus improve performance. We did not see any effect of overfitting for up to two more SFA layers. In the case of overfitting, SFA would be optimized on a specific training trajectory but would perform poorly on test trajectories of the rat through the environment. Overfitting is typically detected out by comparing performance on training and test data sets. As such a performance measure the magnitudes of Δ-values of output units on training and test data was used. In the results presented here, no significant changes of Δ-values occurred. Even the addition of more layers simply reproduced the output of earlier layers. The Δ-values and invariance properties of the SFA layers are depicted in Figure 4.8 for the case of high v_{rel}.

Training set size: More training steps result in a smoother sampling of the virtual reality environment and yield better approximations to the theoretical predictions. We found that a few laps crossing and spanning the whole room within 5.000 training samples were sufficient for the qualitative results already. For too little training data and too few crossings of paths an overfitting effect occurs resulting in a slowly varying activity of the outputs on the training path but not on other (test) paths.

Sparse coding algorithm: As for the choice of the sparse coding algorithm, we found no qualitative difference for different techniques including CuBICA, fastICA, competitive learning, or just finding rotations of the SFA output with maximal kurtosis [Franzius et al., 2007b].

Movement statistics: The choice of movement pattern has a clear impact on the optimal solutions of SFA. The theoretical analysis presented here can in principle predict the solutions for arbitrary movement patterns but for the predictions presented here we made simplifying assumptions to obtain closed form solutions. In spite of these simplifications, the theoretical predictions are still close to the simulation results, e.g., in Section 4.3.1, where the head orientation is restricted to an angular range with respect to the direction of body motion.
In the movement paradigm for the spatial-view cells, the fixated point X changes smoothly over time without abrupt changes. If X instead changed seldom but abruptly, as by saccadic eye movement, similar representations as for smooth changes of X emerge (data not shown), except that the SFA solutions need no longer be similar for adjacent viewpoints. However, in our simulations the similarity of the visual stimuli for adjacent view points often suffices for locally smooth responses.

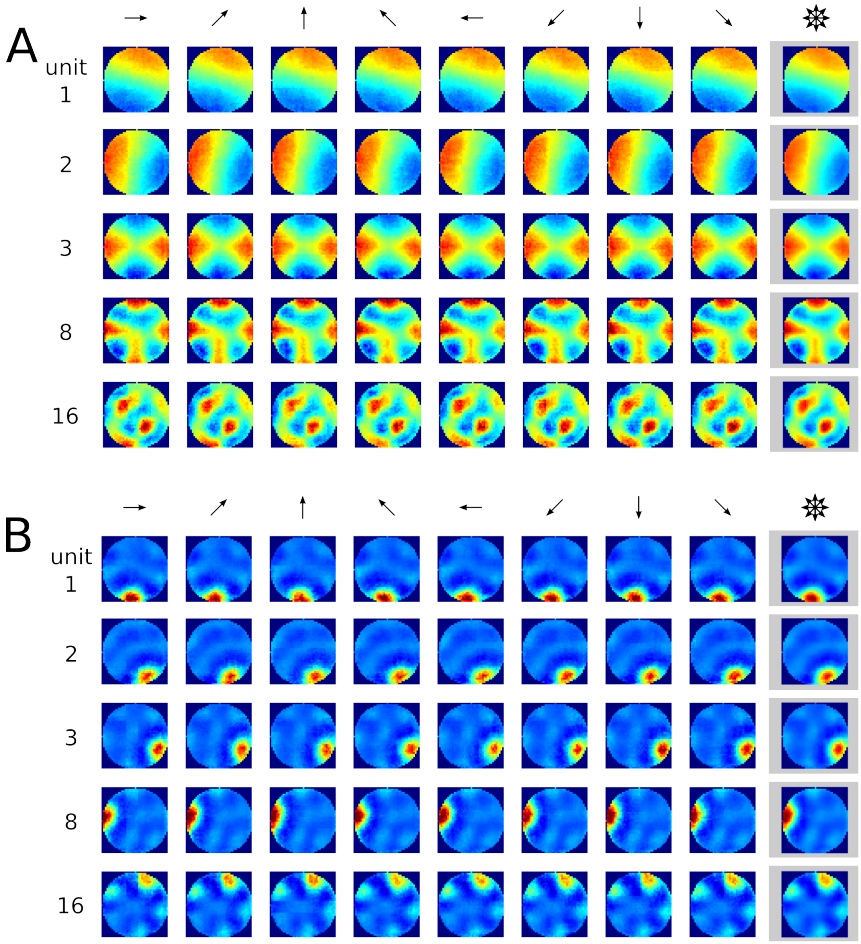

Figure 4.9: Simulation results for a circular room for high relative speed are shown (results with low relative speed are similar to those in Figure 4.3). Each row within each panel shows the response of one unit as a function of position for different head directions (indicated by arrows), as well as the mean value averaged over all head directions (indicated by the superimposed arrows). Blue color denotes low, green intermediate, and red high activity. A: SFA results approximate products of Bessel functions (in radial direction) and harmonic oscillations (in angular directions). B: ICA results of the same simulation.

4.4 Discussion

We have presented a model for the formation of oriospatial cells based on the unsupervised learning principles of slowness and sparseness. The model is feed-forward, instantaneous, and purely sensory driven. The architecture of the model is inspired by the hierarchical organization of the visual system

and applies the identical learning rule, Slow Feature Analysis, on all but the last layer, which performs sparse coding. Our results show that all major oriospatial cell types - place cells, head-direction cells, spatial-view cells, and to some extent even grid cells - can be learned with this approach. We have shown that this model is capable of extracting cognitive information such as an animal's position from complex high-dimensional visual stimuli, which we simulated as views in a virtual environment. The generated representations were coding specifically for some information (e.g., position) and were invariant to others (e.g., orientation). These invariant representations are not explicitly built into the model but induced by the input statistics, which are in turn determined by the room shape and a specific movement paradigm. Nevertheless, the type of learned invariance can be influenced by a temporal adaptation of the learning rate. Control experiments show that the model performance is robust to noise and architectural details. This robustness is also supported by the similarity between simulation results and theoretical predictions for the top SFA level.

Our model comprises sensory processing stages that mimic parts of visual cortex and the hippocampal formation. The model layers cannot be exactly associated with specific brain areas, but we suggest some relations. The behavior of the lower two layers are primarily determined by the visual environment and mostly independent of the spatial movement pattern. In the simulations presented here, we trained the two lower layers only once and only adapted the higher layers for different environments and movement patterns. The first layer could be associated with V1 [Berkes and Wiskott, 2005], the second layer with higher visual areas. Units in the third layer show a periodic non-localized spatial activity pattern (cf. Figure 4A-B), which strongly depends on the movement pattern and might be associated with grid cells in EC. However, two major differences between the SFA representations in the third layer and grid cells are notable. Firstly, grid cells form a hexagonal grid, while the structure in the SFA representations depends on the shape of the room (rectangular rooms yield rectangular SFA patterns). Secondly, the lowest spatial frequency in the SFA representation is half the size of the simulated room, while the peak distances found in EC grid cells show intrinsic spatial scales that range from 39 to 73 cm [Hafting et al., 2005].

The strong influence of room shape on the SFA results is due to the temporally global decorrelation and unit variance constraints in SFA. Thus, SFA requires a decorrelation of activities over arbitrarily long timescales, which might be difficult to achieve in a biologically plausible manner. We expect that a relaxation of these constraints to a limited time window leads to decorrelated representations only within the spatial range that is typically covered by the rat within this time window. This weakens the dependence of the results on the shape of the room and introduces an intrinsic spatial scale as found in EC. Preliminary results indicate that hexagonal activity

patterns can emerge in such a system.

Depending on the movement statistics during learning, representations in the sparse coding layer resemble either place cells as found in hippocampal areas CA1 and CA3 or head-direction cells as found in many areas of the hippocampal formation or spatial-view cells as found in the hippocampal formation of monkeys. For the case of approximately uncorrelated body movement and head direction, the model learns either place or head-direction cells, depending on the relative speed of translation and rotation. For much quicker rotation than translation the model develops orientation-invariant place fields while for much quicker translation than rotation the model develops position-invariant head direction codes. In intermediate cases, e.g., for the linear track, mixed representations such as direction-dependent place fields emerge. Such mixed representations have also been reported in the subicular complex [Cacucci et al., 2004, Sharp, 1996] and medial entorhinal cortex [Sargolini et al., 2006]. In the case of correlated body movement and head direction caused by elongated fixations of objects or positions, the model learns view-specific codes, similar to spatial-view cells in primates.

Although the model is capable of learning place cells and head-direction cells if it learns on distinct adequate movement statistics, a model rat should obviously not have to traverse its environment once with low relative rotational speed to learn head-direction cells and once more with high relative rotational speed to learn place cells. How can both populations be trained with a single given input statistics? For this problem we have considered output from the rat's vestibular system as a possible solution. This system is essential for the oriospatial specificity of head-direction cells and place cells [Stackman and Zugaro, 2005]. Other models like the well established ring attractor model by Skaggs et al. [1995] assume that the head direction system performs angular integration of body motion based on vestibular velocity signals. We hypothesize that these signals could also be used to influence the learning rate of two populations of cells that learn according to our model. One of these populations learns more strongly at periods with high relative translational speed (as signaled by the vestibular angular velocity signals) and the other adapts more strongly for low relative translational speed. The former should develop head-direction cell characteristics and the latter place cell characteristics. In our simulations the model successfully learned both populations with the same input data, one population without learning rate adaptation, and one population with reduced learning rate during quick rotations. Once the model has been trained, the vestibular velocity signal is no longer needed for the model behavior. With learning rate adaptation (LRA) the model neurons effectively learn on a different movement statistics, e.g., head-direction cells learn more strongly at times with relatively high translational speed. Nevertheless, if the real movement statistics contains very few episodes of relatively quick translation at all, the mechanism fails and head-direction cells cannot become position invariant.

The principle of LRA is not limited to changing the effective relative rotational speed, as it can be adapted to reduce learning speed during episodes of quick changes of any feature, as long as some internal signal that is correlated with the change of the feature is available to control the LRA process. We expect that LRA could be used to concurrently learn spatial-view and place cells. This would require a faster change of gaze than in our view-cell simulations above. Then we expect that a population of cells trained without LRA develops place cell characteristics, whereas cells using LRA during episodes of fast fixation point changes develop spatial-view cell characteristics.

Our implementation of the slowness principle involves solving an eigenvalue problem and cannot be considered biologically plausible. However, more plausible implementations exist in the form of gradient-descent learning rules [Hashimoto, 2003, Kayser et al., 2001] and as a spike timing dependent plasticity rule [Sprekeler et al., 2007]. The choice of ICA (and specifically our implementation based on CuBICA) to generate localized representations from nonlocalized codes might seem biologically unrealistic as well [but note Lörincz and Buzsáki, 2000], whereas a formulation in the form of nonlinear Hebbian learning [Oja and Karhunen, 1995] or competitive learning seems more plausible. An in-depth discussion of this topic can be found in [Franzius et al., 2007b].

4.4.1 Related Work

According to Redish's classification, our model is a local view model, for it *"only depends on the local view to explain place cell firing"* [Redish, 1999]. Models of this class usually extract a number of features from sensory inputs in order to obtain a lower-dimensional representation that still carries information about spatial position in the environment but is invariant to everything else. Such models usually do not integrate a path integration system, although for three reasons a complete model of a rodent's hippocampal spatial representation does. Firstly, place fields and head-direction cells are known to fire reliably for several minutes [Muller, 1996] even after removal of most sensory stimuli. A memoryless purely sensory driven system cannot achieve this. Secondly, the almost instantaneous firing of head-direction cells even in new environments and in known environments from positions never seen before requires a system that is operational without any prior sensory input that could be used for learning. Similar arguments apply to the short delay until place cells fire reliably in new environments. We hypothesize that in new environments an instantaneous angular/path integration system dominates until a sensory representation is learned. Thirdly, symmetrical environments lead to symmetrical representations in memoryless models, since there is no way to discriminate identical sensory inputs without a path integration-like memory. However, most place cells in rats

are non-symmetrical in symmetrical environments [Sharp et al., 1990], which underlines the necessity of a path integration system for a complete model. Thus a pure local view model cannot fully explain oriospatial firing properties and therefore many models combine local view and path integration mechanisms [McNaughton et al., 2006, Redish, 1999]. However, pure path integration systems without external sensory input obviously cannot bind to sensory cues like oriospatial cells do (cf. Chapter 3) and thus accumulate integration errors over time. Hence, a sensory coding mechanism, as proposed here, is necessary to complement any such model. In the following, we focus only on local view models.

The model by Wyss et al. [2006] is based on similar principles as our model. It applies a learning rule based on temporal stability to natural stimuli, some of which are obtained from a robot. The resulting spatial representations are localized, resembling hippocampal place fields. The learning rule involves local memory and no explicit sparsification method is applied. The fact that the resulting representations are localized is somewhat surprising, since by itself temporal stability does not lead to localized representations [Franzius et al., 2007b]. The article does not investigate the influence of movement statistics on the learned representations.

The model by Sharp [1991] assumes abstract sensory inputs and acquires a place code by competitive learning, resulting in units that code for views with similar input features. Thus, this model is similar to our model's last layer performing sparsification. Similarly to our results, the degree of head-direction invariance depends on the movement statistics. Unlike our results, however, this is not due to the temporal structure of input views but due to the relative density with which orientation or position are sampled.

The work by Fuhs et al. [1998] uses realistic natural stimuli obtained by a robot and extracts "blobs" of uniform intensity with rectangular or oval shape from these images. Radial basis functions are tuned to blob parameters at specific views, and a competitive learning scheme on these yields place-cell-like representations. Our model agrees with their conclusion that rodents need no explicit object recognition in order to extract spatial information from natural visual stimuli.

The model by Brunel and Trullier [1998] investigates the head-direction dependency of simulated place fields using abstract local views as inputs. A recurrent network learns with an unsupervised Hebbian rule to associate local views with each other, so that their intrinsically directional place cells can become head-direction invariant for maze positions with many rotations. The article also conjectures that movement patterns determine head-direction dependence of place cells, which is consistent with our results.

The results by de Araujo et al. [2001][1] suggest that the size of the rat's

[1] This model was already discussed in Section 3.7.1.

field of view (FOV) is important for the distinction between spatial-view cells and place cells. With a large FOV (as for rats) the animal can see most landmarks from all orientations while an animal with a small FOV (like a monkey) can only see a subset of all landmarks at each time point. We find no dependence of our results on the FOV size for values between 60 and 320 degree as long as the environment is rich enough (i.e., diverse textures, not a single cue card). Instead, our results suggest that differences in the movement statistics play a key role for establishing this difference.

To our knowledge, no prior model allows the learning of place cells, head-direction cells, and spatial-view cells with the same learning rule. Furthermore there are only few models that allow clear theoretical predictions, learn oriospatial cells from (quasi) natural stimuli, and are based on a learning rule that is also known to model early visual processing well.

4.4.2 Future Perspectives

Our simulated visual stimuli come from a virtual reality environment which is completely static during the training of the virtual rat. In this case the slowest features are position, orientation, or view direction as shown before. However, the assumption that the environment remains unchanged during oriospatial cell learning certainly does not hold for the real world. A more realistic environment will include other changing variables like lighting direction, pitch and roll of the head etc. The impact of these variables on the model representations depends on the timescale on which the variables change. For instance, the additional white noise in all SFA layers of the model is ignored since it varies much quicker than position and orientation, but the direction of sunlight might become the slowest feature. Generally, the SFA solutions will depend on any variable whose timescale is equal or slower than the position and orientation of the animal. After the sparse coding step representations will become not only localized in position and/or head direction but in the other variables as well. This behavior is not consistent with the definition of an ideal place or head-direction cell. However, many experiments show correlations of place cell firing with nonspatial variables as well [Redish, 1999]. One particularly interesting instance of such a variable is 'room identity'. If a rat experiences multiple environments, usually transitions between these will occur seldom, i.e., the rat will more often turn and traverse a single room than switch rooms. In this case room identity is encoded by the SFA outputs (data not shown). For n rooms at most $(n-1)$ decorrelated SFA outputs can code for the room identity. The following outputs will then code for a joint representation of space and room identity. After sparse coding, many output units will fire in one room only (the less sparse ones in few rooms), and possibly in a completely unrelated fashion to their spatial firing patterns in another room. This behavior is consistent with the 'remapping' phenomenon in place cells [e.g., Muller and

Kubie, 1987].

A great amount of work has been done investigating the impact of environmental manipulations on oriospatial cell firing in *known* rooms, e.g., shifts and rotations of landmarks relative to each other [Redish, 1999]. How would our model behave after such changes to the learned environment? Such transformations effectively lead to visual input stimuli outside the set of all possible views in the training environment. In this case, we expect the system's performance to deteriorate unless a new representation is learned, but more work is necessary to investigate this question.

Our approach predicts increasing slowness (i.e., decreasing Δ-values of firing rates) in the processing hierarchy between retina and hippocampus. Additionally, place cell and head-direction cell output should be significantly sparser than their inputs. Our main prediction is that changing movement statistics directly influences the invariance properties of oriospatial cells. For instance, an experiment in a linear track where the rat more often turns on mid-track should yield fewer head-direction dependent place cells.

Our model is not limited to processing visual stimuli, as presented here, but can integrate other modalities as well. The integration of olfactory cues, for example, might lead to even more accurate representations and possibly to an independence of the model of visual stimuli (simulated darkness).

Experimentally, the joint positional and orientational dependence of oriospatial cells is hard to measure due to the size of the three-dimensional parameter space, and even more so if the development over time is to be measured. Furthermore, precise data on movement trajectories is rare in the existing literature on oriospatial cells. Accordingly, little data is available to verify or falsify our prediction how the brain's oriospatial codes depend on the movement statistics. As an alternative to determining the movement statistics in behavioral tasks, some work has been done on passive movement of rats, where the movement statistics is completely controlled by the experimenter (e.g., Gavrilov et al. 1998), but these results might not be representative for voluntary motion [Song et al., 2005]. Markus et al. find directional place fields in the center of a plus maze although in the center of the maze more rotations occur than in the arms [Markus et al., 1995]. This could be a contradiction to our model, although not the frequency but the relative speed, which was not measured in [Markus et al., 1995], determines head direction invariance in our model. Overall, the dependence of oriospatial cells on the animal's movement statistics as proposed here remains to be tested experimentally.

4.4.3 Conclusion

We conclude that a purely sensory driven unsupervised system can reproduce many properties of oriospatial cells in the rodent brain, including place cells, head-direction cells, spatial-view cells, and to some extent even grid

cells. These different cell types can be modeled with the same system, and the output characteristics solely depends on the movement statistics of the virtual rat. Furthermore, we showed that the integration of vestibular acceleration information can be used to learn place cells and head-direction cells with the same movement statistics and thus at the same time.

Chapter 5

A Model for Invariant Object Recognition

5.1 Introduction

Sensory signals convey information about the world surrounding us. However, a visual signal can change dramatically even when only a single object is slightly moved or rotated. The visual signal from the retina, for example, varies strongly when distance, position, viewing angle, or lighting conditions change. A high-level representation of object identity in the brain should, however, remain constant or *invariant* under these different conditions. How could the brain extract this abstract information from the highly variable stimuli it perceives? Furthermore, as it is unlikely that the visual brain is completely determined by genetic factors, how can invariant representations be learned from the statistics of the visual stimuli in an unsupervised way? This chapter is based on a cooperation with Niko Wilbert, who worked on the parts about stimulus generation, data analysis, and theoretic predictions.

The primate visual system is organized hierarchically. Object recognition and discrimination are performed in the ventral path that involves the cortical areas V1 (primary visual cortex), V2, V4, and IT (inferotemporal cortex). On the way from V1 to IT, neurons show increasing receptive field sizes, increasing stimulus specificity and invariance. Although massive feedback connections project back down in this hierarchy, feedback across hierarchical layers has only limited impact on simple object recognition tasks. This is because the recognition process is on the order of 100 ms after stimulus onset in macaques and according to [Hung et al., 2005] at least 10 synapses are involved between the retina and IT, leaving only 10 ms on average per synaptic connection.
On the top of this hierarchy, in IT, many neurons are coding for objects with an extreme amount of invariance to position, angle, scale etc. Contrary to the common view of the ventral path as computing only the "what" informa-

tion, though, information about position and size of stimuli is also present in IT [Hung et al., 2005]. Many models of invariant object recognition have been proposed in the last decades [see Rolls and Deco, 2002, for a review]. However, many approaches fail or have not been shown to work for natural stimuli and complex transformations like in-depth rotations. But invariant object recognition is only one task the (primate) brain has to achieve in order to successfully interact with the environment. We do not only need to extract the identity of an object ("What is seen?") independently of its position and view direction, we also want to extract the position of an object ("Where is it?") independently of its identity or viewing angle. Also the relative rotational angle of a viewed object can be crucial ("Does the tiger look at me?"). In principle, we might want a representation of any aspect (i.e., size, viewing angle, lighting direction etc.) independently of all the others and optimally, all these tasks should be solved with a single computational principle. In the following we will refer to the configuration of position and angles relative to the viewer as the *configuration* of an object (sometimes also called a pose). We will call a 2D image of an object in a specific configuration a *view*. In general, the process of extracting the configuration of an object from a view is very hard to solve, especially in the presence of a cluttered background and many different possible objects. In this work we use high-resolution views of complex objects but restrict the problem to the cases of only one object present at any moment and a static homogeneous background. A good model should also *generalize* to previously unseen configurations, i.e., it should learn the relevant statistics of transformations rather than just memorizing specific views. It should also generalize to new objects, e.g., it should successfully estimate the position and orientation[1] of an object that was never shown before.

We apply here a hierarchical model based on the learning principle of *temporal slowness*. This principle has been applied for object recognition before [e.g., Stringer and Rolls, 2002, Einhäuser et al., 2005]. However, our model goes beyond these earlier ones in that it not only extracts translation-invariant and view-invariant representations of object identity but also information about position, and viewing angles. The structure of the resulting representation solely depends on the statistics of the training views. This information is encoded in an analytically predictable way and very simple to decode by linear regression. Except for minor changes (see methods), the model used here is identical to that used earlier for the modeling of place cells and head direction cells in Chapter 4. The complete mathematical framework of the model from [Franzius et al., 2007a] carries over to the problem of invariant object recognition as presented in this chapter.

[1] Generally, there is no canonical "0°"-view of an object, thus a random offset from the absolute angle for a new object is to be expected.

5.1.1 Stimulus Generation

The model was trained and tested with image sequences containing views of different objects. OpenGL was used to render the views of the objects as textured 3D-models in front of a white homogeneous background. To prevent the model from relying on simple color cues (especially for object classification) we only used grayscale views for the results presented here. Two different object classes were used that are described below. For each object class the model was trained with five objects. In the testing phase we added five new objects, which the model had never seen during training.

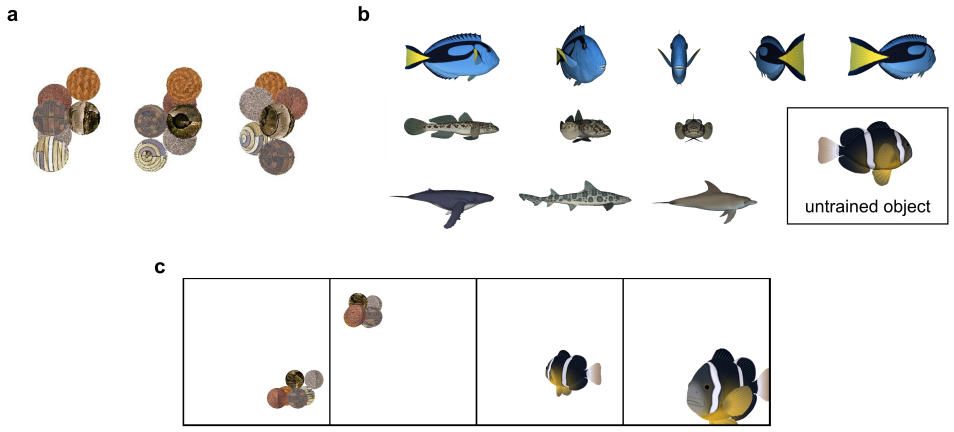

Figure 5.1: Stimuli used for object recognition. A: Sphere objects (each cluster of 6 spheres is one object). The first two views show the same object under different in-depth rotation angles, while the third view shows a different object. **B:** Five fish objects used for training are shown with examples for the effect of in-depth rotation. The fish model on the bottom right is one of the five untrained fish used for testing. **C:** Examples for the training and testing images.

In the first experiment, the objects were clusters of textured spheres as shown in Figure 5.1A and Figure 5.2, which provide the basis for a difficult but generic task. The same six textured spheres (textures from VisTex database [Picard et al., 2002]) were used in different spatial arrangements for all the objects. For each object the spheres were randomly fixed on a hexagonal lattice of size 2x2x3. As the examples in Figure 5.1A illustrate, identifying the rotation angles and identities for such objects is quite difficult for human observers. The choice of identical building blocks for all objects

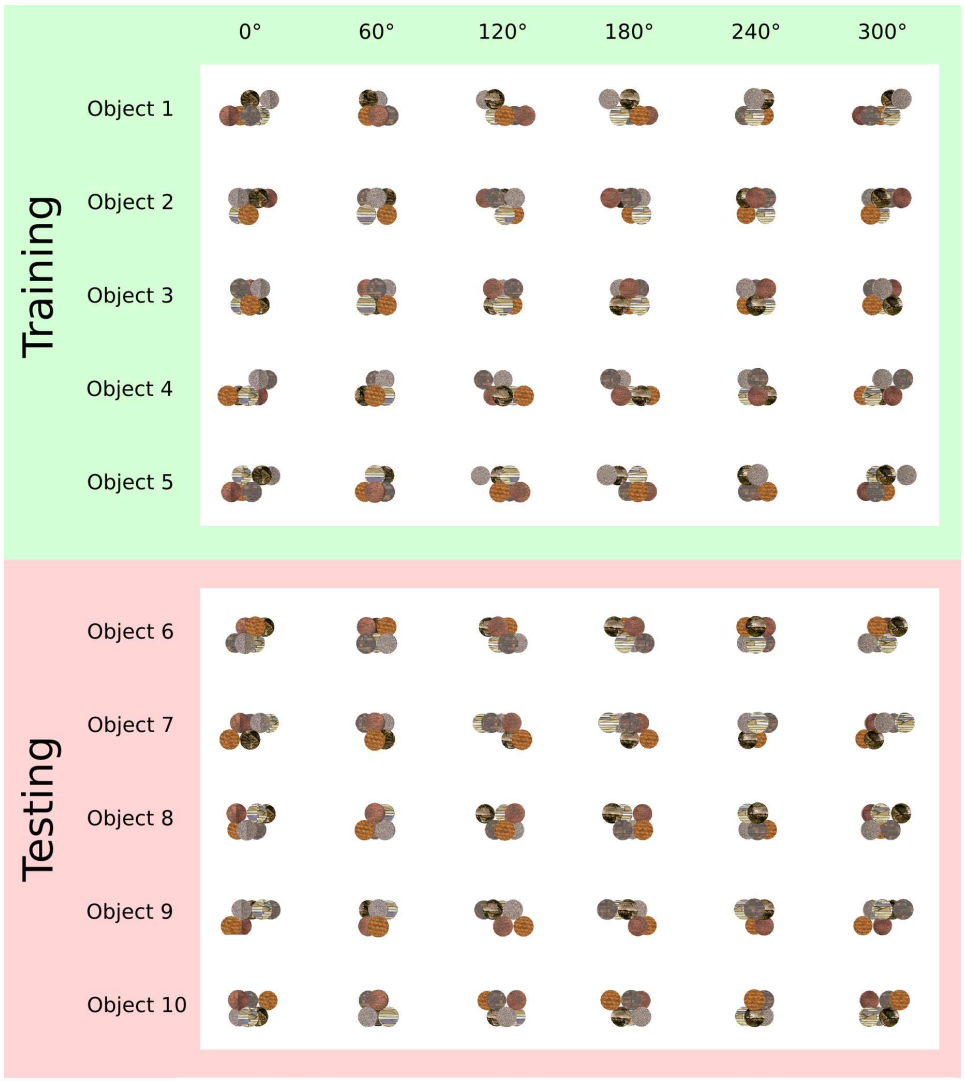

Figure 5.2: Spherical Training and test objects Sphericals. All objects are composed of six identical textured spheres and only differ in the spatial configuration of spheres. Spheres were randomly positioned on a hexagonal 2x2x3 lattice. Each row shows six spatially centered views of a single object with different angles of in-depth rotation. The first five objects were used to train the network and all ten objects were used to test network performance.

is intended to force the model to use high level features for object classification. On the other hand, the common features of the objects help the model to generalize, which is essential for building invariant representations (especially representations that generalize to untrained objects). Using spheres has the advantage that the outline does not give a simple clue for the in-plane rotation angle. In the second experiment, models of different fish (see Figure 5.1b) were used to provide more natural stimuli from a single object class (all models taken from [Toucan Corporation, 2005, with permission]). For sphere objects, the x-coordinate, the y-coordinate (vertical image coordinate), the in-depth rotation angle ϕ_y and the in-plane rotation angle ϕ_z were varied and chosen as configuration variables. x and y range from 0 to 1, ϕ_y and ϕ_z from 0° to 360°. Another configuration variable was the object identity ranging from one to ten, with objects one to five being the training objects. So the transformations the model had to learn consisted of translations in the plane, rotations along the y and z axes (with the in-depth rotation coming first) and changes of object identity. For the fish objects, the configuration variables were x, y, z, ϕ_y and object identity. So compared to the sphere objects we added translations in depth along the z-axis and removed the in-plane rotations. A pure z-translation changes both the object size and the position in the frame, due to the perspective projection. The configurations for the training sequences were generated as a random walk procedure like in Chapter 4 or [Franzius et al., 2007a]. To generate a configuration in the sequence we add a random term to the current spatial and angular velocities of the currently shown object. By adjusting the magnitude of the velocity updates one can effectively choose the timescales of the changes, which are relevant for SFA. The position and angles are then updated according to these velocities. This procedure simulates the effect of inertia and thereby smooths the trajectories. It is then checked if the new configuration violates any boundary conditions (e.g., if the object has left the area of admissible positions in space). If the configuration is not valid it is discarded and new random terms are added to the velocities. Additionally the velocities are decreased (like slowing down if one is facing a wall). The whole procedure produces approximately flat configuration histograms with small deviations at the borders. In each step the object identity was changed with low probability ($p = 0.001$). A blank frame was inserted in between if a switch took place. This procedure adds some realism, as in natural scenes a change in object identity without any changes in other configuration variables generally does not occur.

Theoretical Predictions

The optimal solutions for the infinite-dimensional function space provide very useful predictions for the SFA-output of our model (even though it is based on the finite-dimensional SFA algorithm). Therefore we will now

briefly describe these predictions (for the detailed derivations see [Wiskott, 2003] or [Franzius et al., 2007a]).

The optimal solution for the slowness optimization problem is generally a function of the slowest configuration feature. For example, let us assume that $x_1(t)$ (e.g., the position x or y in Section 5.1.1) is the configuration feature that varies on the slowest timescale (i.e., has the smallest Δ-value). Further assume that the values of x_1 are homogeneously distributed in the interval $[0; L]$ (with $L > 0$) and the distribution of $\dot{x}_1$ is identical for all values of x. Then the slowest output signal $y_1(t) := g_1(\mathbf{x}(t))$ is given by

$$g_1(\mathbf{x}(t)) = \cos\left(\pi \frac{x_1(t)}{L}\right). \tag{5.1}$$

So g_1 is a half cosine over the x_1-interval, i.e., a monotonic code of x_1-position that is invariant to all other configuration parameters. If the timescale of x_2 is only slightly faster than that of x_1, and both are uncorrelated, then the second slowest solution $g_2(\mathbf{x}(t))$ will be a similar half cosine of x_2. The next slowest solutions are of the form

$$\cos\left(\pi \frac{n x_1(t)}{L}\right) \tag{5.2}$$

and

$$\cos\left(\pi \frac{m x_2(t)}{L}\right) \tag{5.3}$$

and products thereof, with $n, m \in \mathbb{N}$, and ordered by slowness. If two such solutions vary on very similar time scale, they can mix resulting in linear combinations of the theoretically predicted solutions.

For angular variables like ϕ_y or ϕ_z, the optimal uncorrelated solutions based on the periodic $\mathbf{x}(t)$ are then $g_1(\mathbf{x}(t)) = \sin(\phi(t))$ and $g_2(\mathbf{x}(t)) = \cos(\phi(t))$. As the rotational angle of a certain stimulus has no natural zero angle, these optimal solution have an arbitrary angular offset for each object. A presentation of different objects in direct succession during the training phase could introduce a common angular offset for all objects but if temporal continuity between the presentation of different objects is disrupted by an intermediate blank stimulus, all representations will have individual random offsets. The same random offset can theoretically also occur for non-periodic variables like position. As position of the stimulus on a white background is a very strong cue, the model probably would require a very large function space in order to assign individual positional codes for each object.

Object identity is a special case as this feature variable takes only discrete[2] values $k(t) \in \{1, 2, \ldots, N\}$. If this configuration variable has the

[2]When using a continuous time t this implies non-differentiable changes of the variable, leading to singularities in the analysis. For practical applications of the SFA-algorithm this does not pose a problem, since time t is discretized in any case.

lowest Δ-value, then the optimal SFA-solution is simply a piecewise constant function. There are $N-1$ uncorrelated optimal solutions, which are all different piecewise constant functions of k [Wiskott and Sejnowski, 2002, Berkes, 2005a].
The blank stimulus between the presentation of different objects forces the SFA outputs to assume individual fixed values during the blank and thus an object-dependence of the outputs does not result in a higher Δ-value. Therefore, for N objects and for each feature like position and orientation, we can expect N additional linear independent outputs which encode a feature for one object and vanish for all other objects.

5.1.2 Network Architecture

The computational model consists of a converging hierarchy of layers of SFA nodes (see Figure 5.3), very similar to the one introduced in Chapter 4. As the main difference in the model architecture, the number of SFA-layers has been increased from three to four because of the increased computational complexity. Additionally, the input now consists of grayscale images of 128 by 128 pixels. The nodes in the lowest layer form a regular (i.e., non-foveated) 24 by 24 grid with partially overlapping receptive fields. The second layer contains 11 by 11 nodes, each receiving input from 4 by 4 layer 1 nodes with neighboring receptive fields, resembling a retinotopical layout. The third layer contains 4 by 4 nodes, each receiving input from 5 by 5 layer 2 nodes with neighboring receptive fields, again in a retinotopical layout. All 4 by 4 layer 3 outputs converge onto a single node in layer 4, whose output we call SFA-output. The output of each node consists of the 32 slowest outputs, except for the top layer where 512 dimensions are used. The layers are trained as in the previous chapter, using 100,000 time points for each layer.

5.1.3 Feature Extraction with Linear Regression

We already presented the form of the expected SFA-output and the problem of linear mixing of features for equal timescales in the section above. The limited function space used by the model will generally lead to deviations from the theoretical predictions, resulting in a smearing of the feature timescales. Thereby a mixing of features in the SFA-output can occur even if the timescales are different, which also affects the higher order harmonics. Overall this makes it practically impossible to prevent the mixing of features for complex applications with many different transformations. Therefore, we did not try to avoid mixing here, but instead we extracted the individual configuration features in a separate post-processing step.

Our main objective here was to show that the relevant features were indeed extracted from the raw image data. The easiest way to do this is

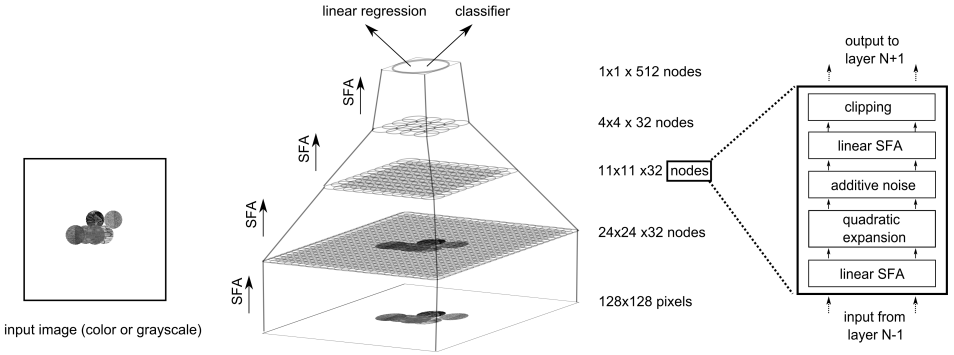

Figure 5.3: Model architecture and stimuli. An input image is fed into the hierarchical network. The circles in each layer denote units, which converge towards the top layer. The same set of steps is applied on each layer, which is visualized on the right hand side.

by calculating a multivariate linear regression of the SFA-output against the known configuration values. This gives us a projection vector for each feature (plus a constant offset value to gauge the result). To calculate the feature value for a given model output one takes the scalar product of the projection vector with the SFA-output vector and adds the offset value.

While the regression procedure is obviously supervised due to the use of the reference configuration values, it nevertheless shows that the relevant signals are easily accessible. Extracting this information from the raw image data linearly is not possible (especially for the angles and object identity, see Section 5.2.3). One should also note, that the dimension of the model output is smaller than the raw image data by two orders of magnitude.

Before actually calculating the regressions, the configuration reference values were binned. For each configuration feature (apart from object identity) 36 bins were used, which is small enough to not influence the results (tests with smaller bin sizes did not show any significant differences). Since the predicted SFA solutions are cosine functions of the position values (see Sections 2.2.2 and 5.1.1), one has to map the reference values correspondingly before calculating the regression. For position, one does not calculate the regression with respect to x, but for the predicted solution $y(x) = \cos(\pi x)$ instead. The result from the linear regression is then mapped back with $\arccos(y)/\pi$ (after clipping the values outside the interval $[-1; 1]$) to get the actual position values.

For those SFA solutions that code for rotational angles, the theory predicts both sine and cosine functions of the angle value (as described in Section 5.1.1). Therefore we calculated regressions with respect to both mappings and then calculated the angles via the arctangent. This automatically matches the global angular offset of the extracted angles to that of

the reference values.

If multiple objects are trained and separated with a blank, the solutions will in general have object dependent angular offsets. We also get additional solutions with complicated angular relationships. These complications rule out global regressions for all objects. The easiest way to avoid this problem is to perform individual regressions for all objects (for each object one regression for sine and one for cosine). While this procedure sacrifices the object invariance it does not affect the translation invariance.

As described in Section 5.1.1, the object identity of N different objects is optimally encoded (under the SFA objective) by up to $N-1$ uncorrelated piecewise constant functions, which are invariant under all other transformations. In the SFA-output this should lead to separated clusters for the trained objects. For new test objects, those SFA-outputs coding for object identity should still be constant but take new values, thereby separating all objects.

To explore the potential classification ability of the model we applied two very simple classifiers on the SFA-output: a k-nearest-neighbor and a Gaussian classifier. The classification performance is affected by imperfect transformation invariance of the identity coding SFA-outputs. This can cause overlap between object clusters. Linear mixing of the identity solutions with other features is also to be expected. Since the linear mixing does not affect cluster separation this should not affect the classifiers. However, the large dimension of the SFA-output has to be matched with enough reference points to capture the shapes of the clusters.

5.2 Results

First, we consider the raw SFA-output for an experiment with a reduced transformation set and clearly separated timescales. Then the experiments with a larger transformation set and similar timescales are analyzed with the methods described above.

5.2.1 Reduced Transformation Set

To illustrate the SFA-outputs of our model we first present a simplified example, with fewer transformations and well separated timescales. This should lead to SFA-outputs which code only one specific feature and are invariant under all other transformations.

Two sphere objects were used for this example, which were switched on a very slow timescale. The other transformations were translations (on a fast timescale) and in-plane rotations (intermediate timescale). Training and testing was done with 150,000 data points and the same model as in the other simulations.

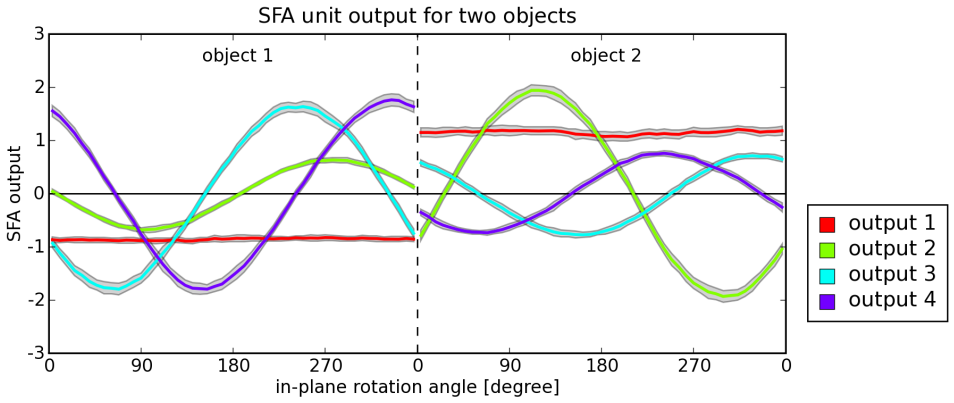

Figure 5.4: Four slowest SFA-outputs for the simulation with reduced transformation set. The in-plane angle dependence of the four slowest SFA-outputs is shown for the two trained sphere objects. The colored lines indicate the mean values, the gray areas cover ± one standard deviation. Note that for each rotational angle the objects were shown at many different positions, so the small deviations illustrate the position invariance of the outputs. The slowest output (red) codes for the object identity. The different amplitudes for the two objects are due to object one appearing slightly more frequent than object two. Outputs three and four code for the sine and cosine of the angle (light and dark blue). The second output (green) is one of the additional solutions taht codes for the rotation angle of a single object and (almost) vanishes for the other object.

As predicted in Section 5.1.1, the slowest output channel codes for object identity (Figure 5.4). Due to poor sampling caused by a low probability of object switches, object one appeared about 9% more often than object two. The more frequently shown object should thus have a lower amplitude for the first component than the other one due to the zero-mean constraint. Outputs two to four illustrate the predicted model outputs coding for the rotational angle (the fourth predicted rotational solution is not shown). Object position is encoded by later SFA-ouputs,a s it corresponds to the fastest fatures (not shown).
In summary, for cases of clearly separated timescales and few simultaneous decorrelated transformations as presented in this section, the model outputs are clearly predictable and directly encode the configuration parameters in an invariant manner.

5.2.2 Full Transformation Set

In this section a larger transformation set is used and the SFA-outputs are subject to supervised postprocessing (see methods). The stimulus set used

for training consisted of 10,000 frames of the five training objects. Data for calculating the regressions and for testing consisted of 100,000 different views of the five objects from the training phase and five new objects (about 10,000 per object), which were generated in the same way as the training data. Half of this data was used to calculate the regressions, the other half for testing.

Position and Rotational Angles

To extract the x and y object coordinates from the model SFA-output we used multivariate linear regression, as described in Section 5.1.3. As one can see in Figure 5.5 and Table 5.1, this works well for the sphere objects, with a standard deviation of 5% for trained objects and 7% for untrained objects. For the trained fish we achieve the same performance as for the trained sphere objects (see Figure 5.6). For untrained fish the additional size variation from the z-transformations take their toll and pull the performance down to 14% for the y-coordinate. This can be improved with an individual regression for each object.

	Spheres		Fish	
	trained	new	trained	new
x	5%	7%	5%	12%
y	5%	7%	5%	14%

Table 5.1: Standard deviations for the coordinate regressions. The values are given in percent relative to the coordinate range. The chance level is 28.9%.

To extract the in-plane and in-depth rotation angles for the spheres, an individual regression for each object was calculated. This still requires invariance of the representations under the other transformations (including the other rotation type). As the results in Table 5.2 show, both rotational angles were extracted with about 15° standard deviation. We also verified that calculating the angles with global regressions for all objects did not work (results for untrained objects were at chance level).

For the fish, the standard deviations are about twice as large as for the spheres, mostly due to systematic errors. The fish models have a very similar front and back view and therefore the model has difficulties to differentiate

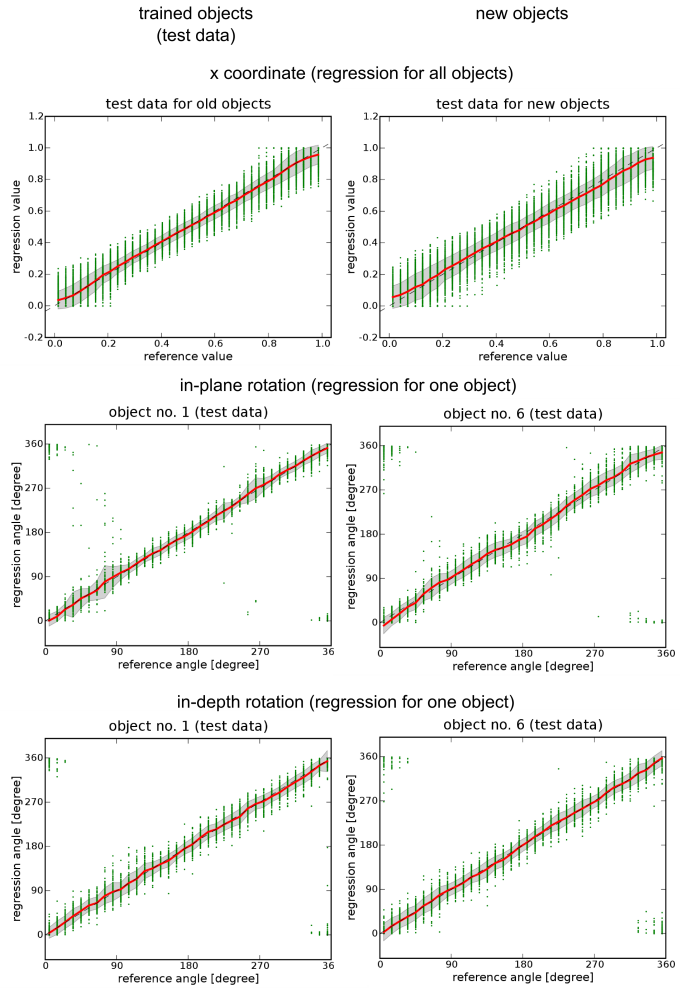

Figure 5.5: Sphere object results for position and angle. The feature values that were calculated with linear regression are plotted against the correct reference values. The green dots are data points, the red line is the mean and the gray area shows ± one standard deviation around the mean value. The regression of the x coordinate was based on all five training objects. For the rotational angles we show object specific regressions (with the untrained object no. 6 on the right).

between those two views, which can be clearly seen in Figure 5.6B. The systematic error introduced by this is mostly responsible for the increase of the standard deviation. When taking the mean absolute error instead (see Figure 5.2), the performance gap to the spheres is smaller because of the reduced influence of outliers.

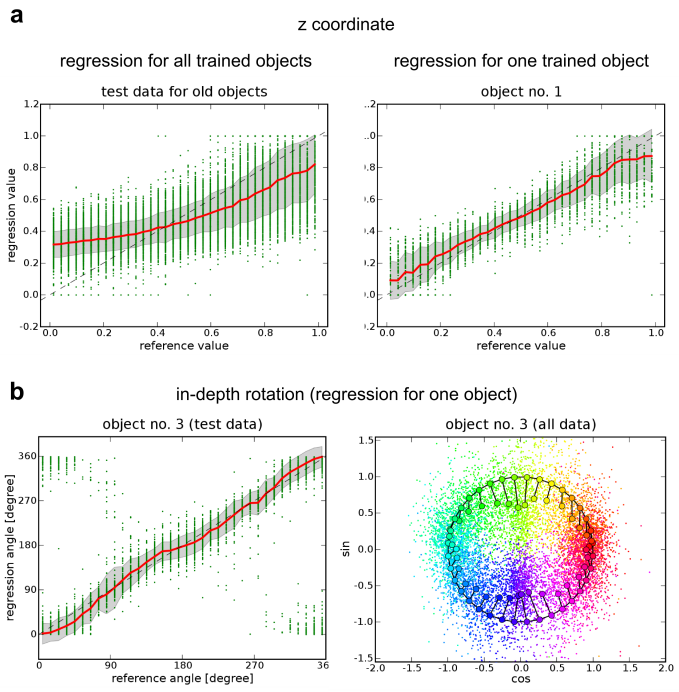

Figure 5.6: Fish object results. **A:** Regression results for the z coordinate. A regression for all training objects leads to large errors (left side). The results improve substantially when an individual regression is performed for each object (see result for object no. 1 on the right). **B:** Regression results for the in-depth rotation angle. The plot on the lower right contains the regression values for the sine and cosine of the in-depth rotation angle. The data points are colored according to the reference angle value. The correct position is indicated by larger points on the black circle, while the inner points are the means for each angle value. The figure eight resemblance of the data cloud is a result of the similar front and back views of the fish models.

		trained						new			mean
	object	1	2	3	4	5	6	7	8	9	10	
Spheres	ϕ_z	14.00°	14.82°	31.56°	11.11°	15.66°	15.92°	10.68°	10.62°	10.34°	13.04°	14.78°
	ϕ_y	13.89°	15.29°	23.57°	16.32°	14.05°	14.45°	13.41°	17.07°	17.09°	13.23°	15.84°
Fish	ϕ_y	13.12°	39.76°	25.49°	51.46°	45.74°	23.61°	45.27°	33.95°	49.98°	35.92°	36.43°
	ϕ_y^*	9.60°	23.31°	16.70°	32.19°	30.28°	16.60°	27.83°	22.18°	32.98°	22.66°	23.43°
	z	0.11	0.09	0.08	0.10	0.09	0.14	0.11	0.11	0.11	0.13	0.11

Table 5.2: Standard deviations for the angles and the z coordinate. The row labeled with ϕ_y^* contains the mean absolute error for ϕ_y, since a large part of the error is systematic due to the 180° pseudo-symmetry (which inflates the standard deviation). The chance level in this case is 104°.

As for the angles, individual regressions for all objects were used to calculate the z coordinate for the fish objects (z-transformations were not used for the sphere objects). This works with a standard deviation of about 11% on average (see Figure 5.2).

Obviously the model has to use the object size to infer the distance, but the area covered by an object also depends on both its identity (due to the different sizes and shapes of the fish) and on the in-depth rotation angle. As one can see in Figure 5.6a, the model has difficulties to compensate for these two factors when a single regression is used for all objects.

Classification

To quantify the classification ability of the model, two classifiers were used on the SFA-output. The classifiers were trained with about 5000 data points per object (i.e., half of the data, as for the regressions). A random fraction of the remaining data points (about 150 per object) was then used to test the classifier performance. The k-nearest-neighbor classifier generally performed with about 96% hit rate (see Table 5.3). As expected, the Gaussian classifier performed not as well, with a hit rate between 88% and 96%. The performance gap between the classifiers and the projection plots in Figure 5.7 suggests that the data clouds for different objects are well separated, but also non-Gaussian to some extent.

5.2.3 Controls

To verify the robustness of our model and justify the model structure we performed several control experiments.

The results presented so far are based on 512 SFA-output channels (i.e., the 512 slowest outputs of the top layer), which may seem excessive. We found that a high channel number increased the quality of the results (see

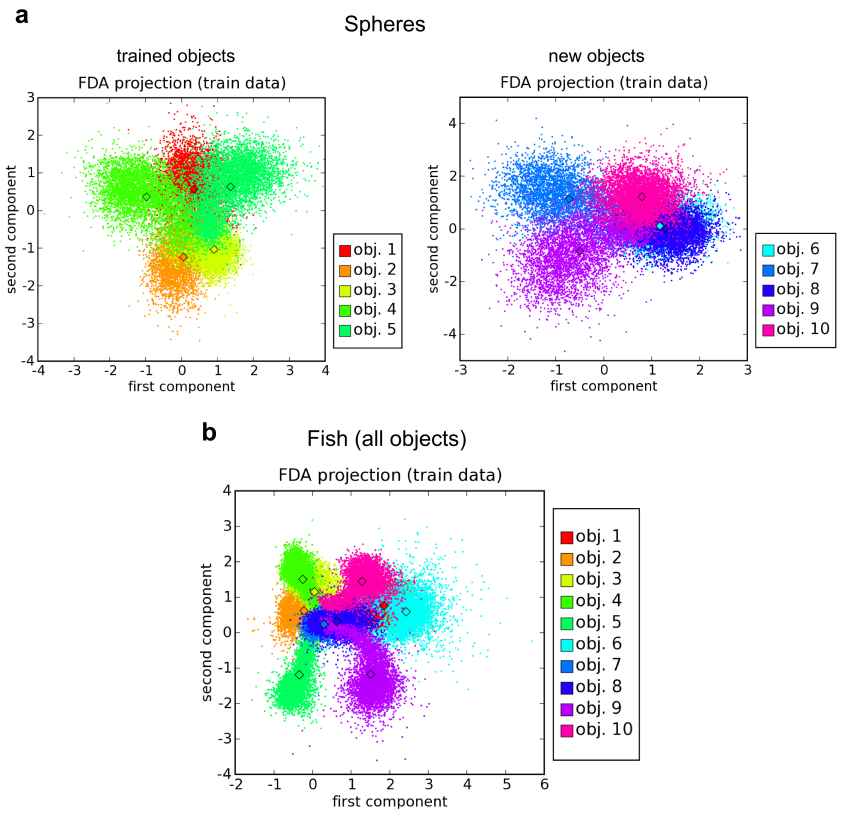

Figure 5.7: 2D projections of the data clusters. The two-dimensional projections of the data points are shown, which are colored according to object identity. The projection plane was chosen to maximize cluster separation (using Fisher discriminant analysis). **A:** Data for trained sphere objects on the left and untrained ones on the right. **B:** Projected data for all ten fish objects.

	trained objects		untrained objects		all objects	
	KNN	G	KNN	G	KNN	G
spheres	96.4	88.3	96.6	95.7	95.0	89.2
fish	97.7	94.6	99.2	88.1	96.7	88.8

Table 5.3: Classifier hit rates in percent. The columns labeled with "KNN" refer to the k-nearest-neighbor classifier ($k = 8$), while those with "G" refer to the Gaussian classifier. Chance level is 10% for all objects and 20% on the sets of training or test objects.

Table 5.4). This holds true for untrained objects, which rules out overfitting effects. The computational cost of working with 512 channels is also quite low (except for the k-nearest neighbor classifier). However, if the number of transformations is small, one can reduce the channel number without

sacrificing performance (see Section 5.2.1).

channels	16	32	64	128	256	512
ϕ_y (std. dev.)	73.0°	58.6°	46.0°	34.7°	24.0°	15.8°
classifier hit rate	23.4	29.5	41.7	59.1	76.0	88.7

Table 5.4: Influence of the number of SFA-output channels (for sphere objects). The table presents results for different channel numbers. The first row shows the standard deviation for ϕ_y for object no. 1. The second row shows the hit rate of the Gaussian classifier on all objects.

To prove that the model can deal with color images (as used in [Franzius et al., 2007a]) we used the colored versions of the fish objects (resulting in a threefold increase in raw data per image). As expected, this slightly increased the classification performance (from about 88.8% hit rate of the Gaussian classifier to about 90% for all ten fish). The standard deviation for the in-depth rotation angle reduced from about 36° down to 19°. On the other hand, the error for z increased slightly (from 11% to 15%).

To verify that the nonlinear expansion is really necessary we trained a purely linear network on the same data (the hierarchical structure was still used, since otherwise the covariance-matrix sizes would have been prohibitive). While the position of the objects was still extracted (with 8% standard deviation), the standard deviation for the angles was about 100° (i.e., close to chance level). The performance of the classifiers was at chance level as well (10.8% hit rate for all objects). We also tested the performance of non-hierarchical linear SFA on view images of reduced size (64×64 pixels). Again the results for the angles were only at chance level, while the classifier performed slightly better (19.4% hit rate). These results demonstrate the necessity of a nonlinear network.

5.2.4 Summary of the Results

The results show that a hierarchical network of SFA units is able to extract identity and configuration information from images of complex objects with complex transformations.

However, our results also show limitations of the model. The dependency between the angles and object identity is a fundamental problem of our approach.

Even more important is the fact that we only used a white background without distractors. In principle the model can deal with changing background, since those changes typically happen on a different timescale than those of object identity. One can also artificially render the model invariant

to the background or distractors by rapidly changing them [Einhäuser et al., 2005], thus sacrificing some realism. Unfortunately our tests in that direction were not very promising. The reason might be due to computational limitations of the model or that for such tasks a simple feed forward network is not sufficient. Attentional top-down mechanism would be a natural extension.

5.3 Discussion

In the previous section we have shown how the hierarchical model learns independent representations of object position, angle, and identity from quasi-natural stimuli. Here, we discuss the relation of our model with existing work and possible extensions.

5.3.1 Related Work

The problem of invariant object recognition has been approached from two sides in the past. One approach is mainly motivated by the "biological implementation" in the (primate) brain with a focus on biological realism, generality and unsupervised learning but was so far mostly limited to very simple stimuli. The other approach comes from computer science, uses sophisticated machine learning approaches, works for complex stimuli, is highly adapted to a specific problem but is often not biologically plausible. In this discussion, we focus on the biologically relevant models.

According to the classifications by Rolls and Deco [2002] and Wiskott [2004], our network belongs to the category of "feature hierarchy based computational object recognition devices". Such systems extract increasingly complex features in a hierarchical system. In contrast to "flat" feature space systems that are typically insensitive to scrambled input images [e.g., SEEMORE: Mel, 1997], object recognition in primates is highly sensitive to the relative location of features, for example position of eyes and the nose in a face [Rolls et al., 1994, Vogels, 1999, Grill-Spector et al., 1998]. The increasing receptive field size in hierarchical feature systems naturally prevents susceptibility to scrambling, as each layer is sensitive to local arrangements of spatial features.

The slowness principle has been applied in many models of the visual system in the past [e.g., Földiák, 1991, Stone and Bray, 1995, Kayser et al., 2001, Wiskott and Sejnowski, 2002, Berkes and Wiskott, 2005, Franzius et al., 2007a]. These unsupervised models learn on naturalistic input data and generate representations similar to neural codes, like V1 simple cells, V1 complex cells, or place cells, head direction cells and spatial view cells in the hippocampal formation.

A similar approach as ours was taken by Wiskott and Sejnowski [2002] for invariant object recognition in a hierarchical system based on SFA. This net-

work is based on a one-dimensional model retina on which one-dimensional stimuli generated from lowpass-filtered white noise were presented. Except for dimensionality, the layer and node structure are very similar to our model. The stimuli were moved across the retina with constant speed. Similarly to our model, some resulting top-level representation encode stimulus identity and are invariant to stimulus position, whereas others encode stimulus position and are invariant to identity. In contrast to our model, the influence of movement statistics on the invariance properties was not established and the complexity of the model in terms of dimensionality, stimulus transformations and realism of stimuli was lower.

VisNet is one of the best-known hierarchical feed-forward neural network models of the primate ventral visual system based on the slowness principle [Rolls and Deco, 2002]. The four model layers are associated with V2, V4, the posterior inferior temporal cortex, and anterior inferior temporal cortex. Each layer consists of a number of computational units with local competition. A number of units in layer $N-1$ project to one unit in layer N in a retinotopic and converging fashion. Similar to our model, receptive field sizes increase from bottom to top layer. Weights in this model are adapted according to the trace rule [Földiák, 1991, Rolls, 1992, Wallis and Rolls, 1997], which is closely related to Slow Feature Analysis [Sprekeler et al., 2007]. The trace rule updates a neuron's input weight vector $\mathbf{w}$ proportionally to the product of the neuron's current input $\mathbf{x}$ and the neuron's *trace value* $\bar{y}^\tau$ according to the rule: $\delta w_j = \alpha \bar{y}^\tau x_j$, where α is the learning rate with $0 \leq \alpha \leq 1$. The trace value $\bar{y}^\tau$ at time step τ is defined as an exponentially weighted neuron's mean activity in the past: $\bar{y}^\tau := (1-\eta)y^\tau + \eta \bar{y}^{\tau-1}$, where $\bar{y}^{\tau-1}$ is the trace value of the prior time step and the magnitude of η defines how much past activities influence the trace. The trace rule becomes identical to Hebbian learning for $\eta = 0$. The weight vector $\mathbf{w}$ has to be normalized explicitly or implicitly in order to prevent unbound growth. Local lateral inhibition is included in the network in order to prevent neighboring nodes from coding for the same features and thus to reduce redundancy. The locality of inhibition allows distant nodes to encode similar features. As in our model, features are learned in an unsupervised manner from the statistics of the network's stimuli. Only on the lowest layer weights are initialized explicitly to Gabor-like structures. Like our model, VisNet can learn a position-invariant (or view-invariant) but object-specific code. In contrast to our model, only a single invariant representation is extracted (i.e., object identity) and the other parameters (e.g., object position, rotation angles, lighting direction) are discarded. In contrast to our model, VisNet has a number of extra variables that need to be tuned per layer: learning rate α, trace length η, inhibition radius σ, contrast δ, and activity percentile a. A further difference comes from the use of sparse coding in VisNet by means of the lateral inhibition and winner-takes-most architecture. In our model, redundancy reduction comes from the decorrelation constraint in SFA, which

is performed in each node at each grid position, but no spatial decorrelation over adjacent nodes in the grid of a layer is enforced. The greatest functional difference between VisNet and our model, however, is our model's ability to code for more than object identity in a structured way, e.g., object-invariant position.
Stringer et al. [2006] use a variant of VisNet with a purely Hebbian learning rule. Interestingly, in experiment 4, the trace rule is applied with a different object presentation statistics. The stimulus presentation was changed such that the two different object types are alternately shown in every second step. The authors find that the resulting representations learned with the trace rule code poorly for object identity. We expect that these representations rather code for object angle.

The model by Einhäuser et al. [2005] applies a very similar learning rule as our model that is basically a gradient descent on the objective function in Section 2.2.2. Similar to VisNet, the first layer uses static Gabor wavelets to model complex cells in V1, whereas the second layer is optimized according to the slowness principle. As this model uses only two layers, the absolute sum of all outputs of the first layer with the same size and orientation is used as input for the second layer, which introduces a hard-coded translation invariance. After optimization, output units represent object identity independently of viewpoint. In contrast to SFA, this approach based on gradient descent might not find the global optima. Between 10 and 50 training objects from the standard COIL database are used as stimuli in front of a homogeneous or cluttered background. The COIL database contains photos of real-world objects centered and rotating on a turntable. Thus object transformations comprise in-depth rotation, changing object identity, some rescaling but no translation, which strongly reduces susceptibility to background clutter. The model incorporates color information (as RGB channels) into the model. In our model, we intentionally reduced the input data to gray scale representations because color histograms alone can often be highly informative about object identity.

The model by Ullman and Bart [2004] is an extension of a simple feature matching system using "extended features". Image regions are learned (called "fragments") that have maximal mutual information between a fragment and the object class it represents. As a fragment might occur at different positions in an image, this model has to explicitly try all possible fragment positions. A fragment might not be visible in all views and so an "extended fragment" consists of a set of alternative fragments (e.g., different views of an eye). The model only addresses a single invariance (i.e., view), which is built-in explicitly by the extended fragment approach.

The model by Rahimi et al. [2005] comes from the computer vision field and has no direct link to a biological system. However, the learning rule applied in this model is very similar to slowness learning. Slowness of the

outputs is enforced by finding a smooth mapping of the inputs to a low-dimensional random field that is governed by a second-order Newtonion process. Few keyframes are labeled by the user and the system maps unseen stimuli to the low-dimensional manifold. This semi-supervised system identifies the complex transformations in high-dimensional video data of people performing complex movements or of talking and grimmassing people.

Berkes [2005a] has performed handwritten digit recognition with a single layer model using quadratic SFA. Here, objects belonging to the same class (e.g., all pictures of a handwritten "0") are presented in random order, whereas changes between classes (e.g., switch of presented zeros to ones) occur seldom (or these switches are discarded entirely from the training sequence). Normalization of inputs with respect to size, orientation, position, and optionally slant was not necessary but improved recognition rate. The most rarely changing feature was object identity. As in our model, up to $n-1$ output features in the form of step functions coded for the identity of the n object classes. The model is computationally efficient and its error rate of 1.4% on test data is better than that of many highly optimized systems. It is interesting to note that, as the time structure of in-class digit presentation is not relevant for this approach, the presentation order is randomized and hence the model becomes equivalent to nonlinear Fisher discriminant analysis.

5.3.2 Outlook and Conclusion

The model proposed here learns invariant object representations based on the statistics of the stimuli presented during the training phase. Specifically, representations of transformations that occur slowly or seldom are formed, while information on other transformations that occur most often or quickly are discarded. An advantage of such a system is that no invariances have to be "hard-wired". The results of the system can be predicted based on theory developed earlier. However, if many transformations occur simultaneously and on similar timescales, solutions tend to mix. For this case a final step of linear regression yields the relevant configuration information.
We show that the system generalizes well to previously unseen configurations (i.e., shows no overfitting for test positions and viewing angles of objects) and to previously unseen objects. However, the system behavior for completely different configurations, like for two simultaneously presented objects, cannot be predicted with our current theory.

Chapter 6

Outlook and Conclusion

This thesis introduced a hierarchical model for unsupervised learning of slowly varying features from artificial but naturalistic high-dimensional input data. The application of the model to spatial learning and to object recognition has demonstrated that such slow features can correspond to highly relevant properties of the environment like an animal's position or orientation in space or the position, identity and rotational angles of a viewed object. For the artificially generated video sequences used here with their well-defined underlying configurations, we have understood the mechanisms governing the solutions of Slow Feature Analysis. Combined with a final step of sparse coding, the model results for spatial application become very similar to representations found in the rodent and primate brain. For these oriospatial cell types, it was shown in Chapter 4 that our model is capable of reproducing most, if not all, known spatial representations in rodents and primates. In contrast to the grids cells found in entorhinal cortex, our simulated grid structures strongly depend on the room shape. A different formulation of the slowness principle with a limited temporal memory might lead to more realistic results in the future.

Three major aspects are considered in the following for a comparison of our model with other models. Firstly, concerning the realism of input data, a model should be able to cope with similar high-dimensional complex stimuli as the brain does. While our artificially generated video sequences are still simpler than real videos (constant illumination, fixed environment/background), they are far more realistic than those of most models, which often rely on distances and angles to point-like landmark. Secondly, the predictive power of a model should be high in order to allow testable predictions and possibly falsification of the model. Our model establishes a clear testable relationship between transformation statistics (i.e., relative movement speed to rotation speed) and model results. Furthermore, the temporal variation of firing rates should decrease in higher hierarchical layers. Thirdly, a model should be general but simple, i.e., it should explain

many phenomena with as little complexity as possible. Although biological systems often lack the simplicity of physical models, following Occam's Razor, a simpler model is better than a more complex model that explains the same data. The same hierarchical model in the same environment was shown to reproduce firing patterns of all major oriospatial cell types only based on different movement statistics. As shown in Chapter 5 the model can also learn invariant object representations based on high-dimensional complex stimuli undergoing realistic transformations. In this aspect, the model presented in this thesis is more general than any other published model. The model is furthermore simple as it is based on the simple and intuitive principles of slowness and sparseness, most parts of the model are built from identical processing units, and a concise mathematical treatment allows to understand the behavior of the model. The flipside of these properties is, however, a higher level of abstraction than that of many other models. Most units in intermediate layers of the hierarchy likely cannot be directly compared with real neurons. Furthermore, the model applies global optimization on real-valued data as opposed to local spike-based computations in the brain. However, approaches for more realistic implementations of the slowness and sparseness principles have been discussed in Chapters 4 and 5.

Both applications for spatial codes and for object invariant representations build on similar low-level representations in the bottom layers of the model. The representations in the lowest layer are known to be good approximations of complex cells in primary visual cortex. Thus the lower model layers could be considered a model of the visual cortical hierarchy, whereas the higher layers also represent different regions, i.e., inferotemporal cortex for object representations and entorhinal cortex and hippocampus for spatial representations. The intermediate model representations have not been thoroughly investigated nor compared to physiological data, for example, from V4 [but note Franzius, 2003]. These representations are likely to strongly depend on the specific choice of receptive field sizes and overlaps, as well as the nonlinearities and the number of layers. It therefore seems reasonable to postpone a comparison of intermediate model representations and higher visual areas to a more biologically realistic implementation of the model.

The hierarchical architecture of the model presented in this thesis has a number of motivations. Firstly, a hierarchical model organization mimics the hierarchical structure of the visual system and possibly other sensory areas. Secondly, although a "flat" single-layer model could yield similar results, the combinatorial explosion of input dimensionality after such an expansion prohibits practical implementations in a computer model as well as in the brain. For the model in Chapter 4, a full flat quadratic expansion of

the inputs in a single layer model, which likely is still insufficient, results in a more than 700,000,000 dimensional representation. Even if the covariance matrix of this representation could be computed in order to solve SFA, most entries will likely be very close to zero, as correlations between distant pixels are very weak [Ruderman and Bialek, 1994]. In contrast to the correlation of single pixels, the "common fate" of object parts spans much longer distances. Object parts will typically not be of homogeneous color but more often of similar texture, so that the pixel values within the different object parts will not have significant correlations but higher-order features like textures will. Thus, although only a small fraction of the possible combinatorial function space is computed in a hierarchical model, long-range correlations of higher-order features can still be captured. The advantages of hierarchical systems are elaborated in more detail in [Rolls and Deco, 2002] and [Hawkins and Blakeslee, 2004].

The model sheds some light on the paradoxical problem of simultaneous invariance and selectivity: how can a unit be invariant to many behaviorally irrelevant transformations of its inputs but at the same time code specifically for a small stimulus class? This thesis proposes a mechanism where a slowness goal function leads to well-understood invariances concerning common transformations, whereas selectivity is enforced by subsequent sparse coding of these invariant representations.

There are a number of shortcomings of the model. Firstly, there is no top-down influence in the model although it is known that the visual cortices have massive feedback from higher to lower areas. We speculate that these feedback connections are critical for attentional mechanisms that have not been integrated into the model. While such an integration would make the analysis of the model behavior much more complicated, an integration of attention seems necessary when, for example, multiple objects are present at the same time.
Secondly, the model provides no form of memory beyond a single time step, although many physiological findings provide clear evidence for a memory based path integration system in the rodent hippocampal formation (see Chapter 3).
Thirdly, the model is implemented using offline-learning. The choice of SFA guarantees to find the global optima efficiently but prohibits direct comparisons with the temporal development of (spatial) codes in the brain. An alternative implementation as an online-learning system based on gradient descent might provide such comparability in the future.
Although we tried to approximate real-world stimuli, the model input is less complex than real video data, especially for the object recognition experiments. For the results in Chapter 5, a homogeneous white background and only one single object at a time was used. Cluttered backgrounds and mul-

tiple objects might require the integration of attentional mechanisms into the model as stated above. Furthermore, as all video data presented here were artificially generated, applicability of the model to real video data has not been shown yet. Although our place cell results were successfully replicated for virtual reality data by an independent group, an implementation for physical robots has proven more complex (personal communication with S. Grünewälder, TU Berlin). In a simulated environment, all configuration parameters are well-known and perfectly controlled, but this is not the case for real robotics, where lighting conditions and room configurations can change. If such changes occur on a slow timescale, our model would encode such features, which can be very hard to identify. It remains to be shown if our results can be reproduced ”in the real world”.

Bibliography

D. Amaral and P. Lavenex. Hippocampal neuroanatomy. In *The hippocampus book*, pages 37–114. Oxford University Press, 2007.

D. G. Amaral and M. P. Witter. Hippocampal formation. In G. Paxinos, editor, *The Rat Nervous System*, pages 443–493. Academic Press, San Diego, USA, second edition, 1995.

P. Andersen, R. Morris, D. Amaral, T. Bliss, and J. O'Keefe, editors. *The Hippocampus Book*. Oxford University Press, 2007.

J. C. Baird, J. S. Taube, and D. V. Peterson. Statistical and information properties of head direction cells. *Perception & psychophysics*, 63(6):1026–1037, 2001.

C. Barry, R. Hayman, N. Burgess, and K. J. Jeffery. Experience-dependent rescaling of entorhinal grids. *Nature Neuroscience*, 10:682–684, 2007. doi: 10.1038/nn1905.

E. B. Baum, J. Moody, and F. Wilczek. Internal representations for associative memory. *Biological Cybernetics*, 59(4–5):217–228, 1988.

S. Becker and R. S. Zemel. Unsupervised learning with global objective functions. In *The Handbook of Brain Theory and Neural Networks*, pages 1183–1187. MIT Press, 2003.

T. Bell and T. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.

S. Benhamou. An analysis of movements of the wood mouse apodemus sylvaticus in its home range. *Behavioral Processes*, 22:235–250, 1990.

P. Berkes. Handwritten digit recognition with nonlinear fisher discriminant analysis. *Proceedings of ICANN 2005*, 2(LNCS 3696):285–287, 2005a.

P. Berkes. *Temporal slowness as an unsupervised learning principle*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät I, Universitätsbibliothek, 2005b.

P. Berkes and L. Wiskott. Slow feature analysis yields a rich repertoire of complex cell properties. *Journal of Vision*, 5(6):579–602, 2005. http://journalofvision.org/5/6/9/, doi:10.1167/5.6.9.

P. Berkes and T. Zito. Modular toolkit for data processing (version 2.0). `http://mdp-toolkit.sourceforge.net`, 2005.

T. Blaschke. *Independent Component Analysis and Slow Feature Analysis*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät I, Universitätsbibliothek, 2005.

T. Blaschke and L. Wiskott. CuBICA: Independent component analysis by simultaneous third- and fourth-order cumulant diagonalization. *IEEE Transactions on Signal Processing*, 52(5):1250–1256, 2004.

T. Blaschke, P. Berkes, and L. Wiskott. What is the relationship between slow feature analysis and independent component analysis? *Neural Computation*, 18(10):2495–2508, 2006.

A. Bray and D. Martinez. Kernel-based extraction of slow features: Complex cells learn disparity and translation invariance from natural images. In *NIPS: Neural Information Processing*, volume 15, pages 253–260, 2002.

J. E. Brown, B. J. Yates, and J. S. Taube. Does the vestibular system contribute to head direction cell activity in the rat? *Physiology & behavior*, 77(4-5):743–748, 2002.

N. Brunel and O. Trullier. Plasticity of directional place fields in a model of rodent CA3. *Hippocampus*, 8:651–665, 1998.

Y. Burak, T. Brookings, and I. Fiete. Triangular lattice neurons may implement an advanced numeral system to precisely encode rat position over large ranges, 2006. URL `http://www.citebase.org/abstract?id=oai:arXiv.org:q-bio/0606005`.

N. Burgess and J. O'Keefe. Hippocampus: Spatial models. In P. H. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 539–543. MIT Press, 2003.

N. Burgess, F. Cacucci, C. Lever, and J. O'Keefe. Characterizing multiple independent behavioral correlates of cell firing in freely moving animals. *Hippocampus*, 15(2):149–153, 2005.

G. Buzsáki. Large-scale recording of neuronal ensembles. *Nature Neuroscience*, 7(5):446–451, 2004.

F. Cacucci, C. Lever, T. J. Wills, N. Burgess, and J. O'Keefe. Theta-modulated place-by-direction cells in the hippocampal formation in the rat. *Journal of Neuroscience*, 24(38):8265–8277, 2004.

J. L. Calton and J. S. Taube. Degradation of head direction cell activity during inverted locomotion. *Journal of Neuroscience*, 25(9):2420–2428, 2005.

J. L. Calton, R. W. Stackman, J. P. Goodridge, W. B. Archey, P. A. Dudchenko, and J. S. Taube. Hippocampal place cell instability after lesions of the head direction cell network. *Journal of Neuroscience*, 23(30):9719–9731, 2003.

M. K. Chawla, J. F. Guzowski, V. Ramirez-Amaya, P. Lipa, K. L. Hoffman, L. K. Marriott, P. F. Worley, B. L. McNaughton, and C. A. Barnes. Sparse, environmentally selective expression of arc RNA in the upper blade of the rodent fascia dentata by brief spatial experience. *Hippocampus*, 15(5):579–586, 2005.

F. Creutzig and H. Sprekeler. Predictive coding and the slowness principle: An information-theoretic approach. *Neural Computation*, accepted, 2008.

A. Czurkó, H. Hirase, J. Csicsvari, and G. Buzsáki. Sustained activation of hippocampal pyramidal cells by 'space clamping' in a running wheel. *European Journal of Neuroscience*, 11(1):344–352, 1999.

I. E. T. de Araujo, E. T. Rolls, and S. M. Stringer. A view model which accounts for the spatial fields of hippocampal primate spatial view cells and rat place cells. *Hippocampus*, 11:699–706, 2001.

M. E. Deutschlander, J. B. Phillips, and S. C. Borland. The case for light-dependent magnetic orientation in animals. *Journal of Experimental Biology*, 202(8):891–908, 1999.

H. Eichenbaum and N. J. Cohen. Representation in the hippocampus: what do hippocampal neurons encode? *Trends in Neuroscience*, 11(6):244–248, 1988.

H. Eichenbaum, M. Kuperstein, A. Fagan, and J. Nagode. Cue-sampling and goal-approach correlates of hippocampal unit activity in rats performing an odor-discrimination task. *Journal of Neuroscience*, 7(3):716–732, 1987.

H. Eichenbaum, P. Dudchenko, E. Wood, M. Shapiro, and H. Tanila. The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron*, 23:209–226, 1999.

W. Einhäuser, J. Hipp, J. Eggert, E. Körner, and P. König. Learning viewpoint invariant object representations using a temporal coherence principle. *Biological Cybernetics*, 93:79–90, 2005.

A. D. Ekstrom, M. J. Kahana, J. B. Caplan, T. A. Fields, E. A. Isham, E. L. Newman, and I. Fried. Cellular networks underlying human spatial navigation. *Nature*, 425:184–188, 2003.

A. S. Etienne. The control of short-distance homing in the golden hamster. In P. Ellen and C. Thinus-Blanc, editors, *Cognitive Processes and Spatial Orientation in Animals and Man, volume I*, Experimental Animal Psychology and Ethology, pages 233–251. Martinus Nijhoff Publishers, Boston, 1987.

A. A. Fenton. Where am I? *Science*, 315(5814):947–949, 2007.

D. J. Field. What is the goal of sensory coding? *Neural Computation*, 6: 559–601, 1994.

P. Földiák. Learning invariance from transformation sequences. *Neural Computation*, 3:194–200, 1991.

P. Földiák and M. Young. Sparse coding in the primate cortex. In *The Handbook of Brain Theory and Neural Networks*, pages 1064–1068. MIT Press, 2003.

T. C. Foster, C. A. Castro, and B. L. McNaughton. Spatial selectivity of rat hippocampal neurons: dependence on preparedness for movement. *Science*, 244(4912):1580–1582, 1989.

S. E. Fox and J. B. Ranck Jr. Electrophysiological characteristics of hippocampal complex-spike cells and theta cells. *Experimental Brain Research*, 41:399–410, 1981.

L. M. Frank, E. N. Brown, and M. Wilson. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, 27:169–178, 2000.

L. M. Frank, G. B. Stanley, and E. N. Brown. Hippocampal plasticity across multiple days of exposure to novel environments. *Journal of Neuroscience*, 24(35):7681–7689, 2004.

M. Franzius. Unüberwachtes Lernen von Texturen in einem hierarchischen Neuronalen Netzwerk mittels natürlicher Stimuli. Diplomarbeit, Lehrstuhl für Grafische Systeme, Brandenburgische Technische Universität Cottbus, 2003.

M. Franzius, H. Sprekeler, and L. Wiskott. Slowness and sparseness lead to place, head-direction and spatial-view cells. *Public Library of Science (PLoS) Computational Biology*, 3(8):e166, 2007a. doi: 10.1371/journal.pcbi.0030166.

M. Franzius, R. Vollgraf, and L. Wiskott. From grids to places. *Journal of Computational Neuroscience*, 22(3), 2007b.

M. Franzius, N. Wilbert, and L. Wiskott. Object recognition with the slowness principle. *(in preparation)*, 2007c.

M. C. Fuhs and D. S. Touretzky. A spin glass model of path integration in rat medial entorhinal cortex. *Journal of Neuroscience*, 26(16):4266–4276, 2006.

M. C. Fuhs, A. D. Redish, and D. S. Touretzky. A visually driven hippocampal place cell model. *Proceedings of the Sixth Annual Conference on Computational Neuroscience: Trends in Research*, pages 101–106, 1998.

K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, 1980.

M. Fyhn, S. Molden, M. P. Witter, E. I. Moser, and M.-B. Moser. Spatial representation in the entorhinal cortex. *Science*, 305(5688):1258–1264, 2004.

V. V. Gavrilov, S. I. Wiener, and A. Berthoz. Discharge correlates of hippocampal complex spike neurons in behaving rats passively displaced on a mobile robot. *Hippocampus*, 8(5):475–490, 1998.

G. Genaro and W. R. Schmidek. Exploratory activity of rats in three different environments. *Ethology*, 106(9):849–859, 2000.

P. Georges-François, E. T. Rolls, and R. G. Robertson. Spatial view cells in the primate hippocampus: allocentric view not head direction or eye position or place. *Cerebral Cortex*, 9(3):197–212, 1999.

E. J. Golob and J. S. Taube. Head direction cells in rats with hippocampal or overlying neocortical lesions: Evidence for impaired angular path integration. *Journal of Neuroscience*, 19(16):7198–7211, 1999.

E. J. Golob and J. S. Taube. Head direction cells and episodic spatial information in rats without a hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, 94(14):7645–7650, 1997.

J. P. Goodridge, P. A. Dudchenko, K. A. Worboys, E. J. Golob, and J. S. Taube. Cue control and head direction cells. *Behavioral Neuroscience*, 112(4):749–761, 1998.

K. M. Gothard, W. E. Skaggs, and B. L. McNaughton. Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues. *Journal of Neuroscience*, 16(24):8027–8040, 1996.

D. J. Graham and D. J. Field. Sparse coding in the neocortex. In J. H. Kaas and L. A. Krubitzer, editors, *Evolution of the Nervous Systems*. (in press), 2007.

B. Greenstein and A. Greenstein. *Color Atlas of Neuroscience: Neuroanatomy and Neurophysiology*. Thieme Medical Publishers, 2000.

K. Grill-Spector, T. Kushnir, T. Hendler, S. Edelman, Y. Itzchak, and R. Malach. A sequence of object-processing stages revealed by fMRI in human occiptal lobe. *Human Brain Mapping*, 6:316–328, 1998.

C. G. Gross. Genealogy of the "grandmother cell". *Neuroscientist*, 8(5): 512–518, 2002.

A. Guanella and P. F. Verschure. A model of grid cells based on a path integration mechanism. In *Proceedings of Artificial Neural Networks – ICANN 2006. Lecture Notes in Computer Science.*, volume 4131, pages 740–749, Berlin, 2006. Springer.

T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005. doi:10.1038/nature03721.

E. Hargreaves, G. Rao, I. Lee, and J. J. Knierim. Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science*, 308: 1792–1794, 2005.

W. Hashimoto. Quadratic forms in natural images. *Network: Computation in Neural Systems*, 14(4):765–788, 2003.

J. Hawkins and S. Blakeslee. *On Intelligence*. Times Books, 2004.

A. J. Hill. First occurrence of hippocampal spatial firing in a new environment. *Experimental Neurology*, 62(2):282–97, 1978.

S. A. Hollup, S. Molden, J. G. Donnett, M.-B. Moser, and E. I. Moser. Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *Journal of Neuroscience*, 21(5):1635–1644, 2001.

C. Hölscher, A. Schnee, H. Dahmen, L. Setia, and H. A. Mallot. Rats are able to navigate in virtual environments. *Journal of Experimental Biology*, 208:561–569, 2005.

S. L. Hopp and W. Timberlake. Odor cue determinants of urine marking in male rats (rattus norvegicus). *Behavioral Neural Biology*, 37(1):162–72, 1983.

E. Hori, Y. Nishio, K. Kazui, K. Umeno, E. Tabuchi, K. Sasaki, S. Endo, T. Ono, and H. Nishijo. Place-related neural responses in the monkey hippocampal formation in a virtual space. *Hippocampus*, 15(8):991–996, 2005.

A. Hughes. A schematic eye for the rat. *Vision Research*, 19:569–588, 1978.

C. P. Hung, G. Kreiman, T. Poggio, and J. J. DiCarlo. Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310:863–866, 2005.

J. Hurri and A. Hyvärinen. Simple-cell-like receptive fields maximize temporal coherence in natural video. *Neural Computation*, 15(3):663–691, 2003.

A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10:626–634, 1999a.

A. Hyvärinen. Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 11:1739–1768, 1999b.

A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999c.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley Interscience, New York, 2001.

N. Intrator. Competitive learning. In *The Handbook of Brain Theory and Neural Networks*, pages 238–241. MIT Press, 2003.

I. T. Jolliffe. *Principal Component Analysis*. Springer, New York, 2002.

M. W. Jung and B. L. McNaughton. Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus*, 3(2):165–182, 1993.

M. W. Jung, S. I. Wiener, and B. L. McNaughton. Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *Journal of Neuroscience*, 14:7347–7356, 1994.

C. Kayser, W. Einhäuser, O. Dümmer, P. König, and K. Körding. Extracting slow subspaces from natural videos leads to complex cells. *Artificial Neural Networks - ICANN 2001 Proceedings*, pages 1075–1080, 2001.

J. J. Knierim. Neural representations of location outside hippocampus. *Learning & Memory*, 13:405–415, 2006.

J. J. Knierim, H. S. Kudrimoti, and B. L. McNaughton. Place cells, head direction cells, and the learning of landmark stability. *Journal of Neuroscience*, 15(3 Pt 1):1648–1659, 1995.

K. P. Körding, C. Kayser, W. Einhäuser, and P. König. How are complex cell properties adapted to the statistics of natural stimuli? *Journal of Neurophysiology*, 91(1):206–212, 2004.

A. K. Lee, I. D. Manns, B. Sakmann, and M. Brecht. Whole-cell recordings in freely moving rats. *Neuron*, 51(4):399–407, 2006.

P. Lennie. The cost of cortical computation. *Current Biology*, 13:493–497, 2003.

S. Leutgeb, J. K. Leutgeb, C. A. Barnes, E. I. Moser, B. L. McNaughton, and M.-B. Moser. Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, 309:619–623, 2005.

A. Lörincz and G. Buzsáki. Two-phase computational model training long-term memories in the entorhinal-hippocampal region. *Annals of the New York Academy of Sciences*, 911:83–111, 2000.

N. Ludvig, H. M. Tang, B. C. Gohil, and J. M. Botero. Detecting location-specific neuronal firing rate increases in the hippocampus of freely-moving monkeys. *Brain Research*, 1014(1–2):97–109, 2004.

E. J. Markus, Y. L. Qin, B. Leonard, W. E. Skaggs, B. L. McNaughton, and C. A. Barnes. Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *Journal of Neuroscience*, 15(11):7079–7094, 1995.

A. P. Maurer, S. R. VanRhoads, G. R. Sutherland, P. Lipa, and B. L. McNaughton. Self-motion and the origin of differential spatial scaling along the septo-temporal axis of the hippocampus. *Hippocampus*, 15:841–852, 2005.

J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, 1995.

B. L. McNaughton, C. A. Barnes, and J. O'Keefe. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Experimental Brain Research*, 52:41–49, 1983a.

B. L. McNaughton, J. O'Keefe, and C. A. Barnes. The stereotrode: a new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. *Journal of Neuroscience Methods*, 8(4):391–397, 1983b.

B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, and M.-B. Moser. Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, 7:663–678, 2006.

B. W. Mel. SEEMORE: Combining color, shape and texture histogramming in a neurally-inspired approach to visual object recognition. *Neural Computation*, 9(4):777–804, 1997.

V. M. Miller and P. J. Best. Spatial correlates of hippocampal unit activity are altered by lesions of the fornix and entorhinal cortex. *Brain Research*, 194:311–323, 1980.

G. Mitchison. Removing time variation with the anti-Hebbian differential synapse. *Neural Computation*, 3:312–320, 1991.

M. L. Mittelstaedt and H. Mittelstaedt. Homing by path integration in a mammal. *Naturwissenschaften*, 67:566–567, 1980.

E. Moser and M.-B. Moser. Grid cells. *Scholarpedia*, page 16073, 2007.

E. I. Moser and O. Paulsen. New excitement in cognitive space: between place cells and spatial memory. *Current Opinion in Neurobiology*, 11: 745–751, 2001.

M.-B. Moser and E. I. Moser. Functional differentiation in the hippocampus. *Hippocampus*, 6(8):608–619, 1998.

R. Muller. A quarter of a century of place cells. *Neuron*, 17:813–822, 1996.

R. U. Muller and J. L. Kubie. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience*, 7(7):1951–1968, 1987.

R. U. Muller, J. L. Kubie, and J. B. Ranck Jr. Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *Journal of Neuroscience*, 7:1935–1950, 1987.

R. U. Muller, E. Bostock, J. S. Taube, and J. L. Kubie. On the directional firing properties of hippocampal place cells. *Journal of Neuroscience*, 14 (12):7235–7251, 1994.

K. Nakazawa, T. J. McHugh, M. A. Wilson, and S. Tonegawa. NMDA receptors, place cells and hippocampal spatial memory. *Nature Reviews Neuroscience*, 5(5):361–72, 2004.

E. Oja. A simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, 15:267–273, 1982.

E. Oja. Principal components, minor components, and linear neural networks. *Neural Networks*, 5(6):927–935, 1992.

E. Oja and J. Karhunen. Signal separation by nonlinear Hebbian learning. *Computational Intelligence: A Dynamic System Perspective*, pages 83–97, 1995.

J. O'Keefe. Hippocampal neurophysiology in the behaving animal. In *The hippocampus book*, pages 475–548. Oxford University Press, 2007.

J. O'Keefe. Do hippocampal pyramidal cells signal non-spatial as well as spatial information? *Hippocampus*, 9:352–364, 1999.

J. O'Keefe and N. Burgess. Geometric determinants of the place fields of hippocampal neurons. *Nature*, 381:425–428, 1996.

J. O'Keefe and N. Burgess. Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus*, 15:853–866, 2005.

J. O'Keefe and D. H. Conway. Hippocampal place units in the freely moving rat: why they fire where they fire. *Experimental Brain Research*, 31(4): 573–90, 1978.

J. O'Keefe and J. Dostrovsky. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat. *Brain Research*, 34:171–5, 1971.

J. O'Keefe and L. Nadel. *The hippocampus as a cognitive map*. Oxford University Press, Oxford, UK, 1978.

J. O'Keefe and M. L. Recce. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, 3(3):317–330, 1993.

J. O'Keefe and A. Speakman. Single unit activity in the rat hippocampus during a spatial memory task. *Experimental Brain Research*, 68:1–27, 1987.

H. G. Olbrich and H. Braak. Ratio of pyramidal cells versus non-pyramidal cells in sector CA1 of the human ammon's horn. *Anatomy and Embryology*, 173(1):105–110, 1985.

B. A. Olshausen. Principles of image representations in visual cortex. In L. M. Chalupa and J. S. Werner, editors, *The visual neuroscienes*, pages 1603–1615. MIT Press, 2003.

B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381: 607–609, 1996.

B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–3325, 1997.

B. A. Olshausen and D. J. Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14:481–487, 2004.

S. M. O'Mara, E. T. Rolls, A. Berthoz, and R. P. Kesner. Neurons responding to whole-body motion in the primate hippocampus. *Journal of Neuroscience*, 14(11 Pt 1):6511–6523, 1994.

T. Ono, K. Nakamura, H. Nishijo, and S. Eifuku. Monkey hippocampal neurons related to spatial and nonspatial functions. *Journal of Neurophysiology*, 70:1516–1529, 1993.

T. Otto and H. Eichenbaum. Neuronal activity in the hippocampus during delayed non-match to sample performance in rats: evidence for hippocampal processing in recognition memory. *Hippocampus*, 2(3):323–334, 1992.

R. Picard, C. Graczyk, S. Mann, J. Wachman, L. Picard, and L. Campbell. Vision texture. Downloaded from `http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html`, 2002.

B. Poucet, P. P. Lenck-Santini, V. E. Paz-Villagrán, and E. Save. Place cells, neocortex and spatial navigation: a short review. *Journal of physiology (Paris)*, 97:537–546, 2003.

G. J. Quirk, R. U. Muller, and J. L. Kubie. The firing of hippocampal place cells in the dark depends on the rat's recent experience. *Journal of Neuroscience*, 10(6):2008–2017, 1990.

G. J. Quirk, R. U. Muller, J. L. Kubie, and J. B. Ranck Jr. The positional firing properties of medial entorhinal neurons: description and comparison with hippocampal place cells. *Journal of Neuroscience*, 12(5):1945–1963, 1992.

A. Rahimi, B. Recht, and T. Darrell. Learning appearance manifolds from video. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, pages 868–875, Washington, DC, USA, 2005. IEEE Computer Society. doi: http://dx.doi.org/10.1109/CVPR.2005.204.

J. B. Ranck Jr. Studies on single neurons in dorsal hippocampal formation and septum in unrestrained rats. I. behavioral correlates and firing repertoires. *Experimental Neurology*, 41(2):461–531, 1973.

J. B. Ranck Jr. Head direction cells in the deep cell layer of dorsal presubiculum in freely moving rats. In G. Buzsáki and C. H. Vanderwolf, editors, *Electrical activity of archicortex*, pages 217–220. Akadémiai Kiadó, Budapest, 1985.

P. R. Rapp and M. Gallagher. Preserved neuron number in the hippocampus of aged rats with spatial learning deficits. *Proceedings of the National Academy of Sciences of the United States of America*, 93:9926–9930, 1996.

A. D. Redish. *Beyond the Cognitive Map - From Place Cells to Episodic Memory*. MIT Press, 1999.

A. D. Redish. The hippocampal debate: are we asking the right questions? *Behavioral Brain Research*, 127(1–2):81–98, 2001.

T. E. Robinson. Hippocampal rhythmic slow activity (RSA; theta): a critical analysis of selected studies and discussion of possible species-differences. *Brain Research*, 203(1):69–101, 1980.

E. T. Rolls. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions of the Royal Society London*, 335:11–21, 1992.

E. T. Rolls. Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus*, 9:467–480, 1999.

E. T. Rolls. Neurophysiological and computational analyses of the primate presubiculum, subiculum and related areas. *Behavioral Brain Research*, 174:289–303, 2006.

E. T. Rolls. The representation of information about faces in the temporal and frontal lobes. *Neuropsychologia*, 45:124–143, 2007.

E. T. Rolls and G. Deco. *The Computational Neuroscience of Vision*. Oxford University Press, New York, 2002.

E. T. Rolls and S. M. Stringer. Invariant visual object recognition: A model, with lighting invariance. *Journal of Physiology - Paris*, 100:43–62, 2006.

E. T. Rolls, M. J. Tovee, D. G. Purcell, A. L. Stewart, and P. Azzopardi. The responses of neurons in the temporal cortex of primates, and face identification and detection. *Experimental Brain Research*, 101(3):473–484, 1994.

E. T. Rolls, R. G. Robertson, and P. Georges-François. Spatial view cells in the primate hippocampus. *European Journal of Neuroscience*, 9:1789–1794, 1997a.

E. T. Rolls, A. Treves, R. G. Robertson, P. Georges-Fancois, and S. Panzeri. Information about spatial view in an ensemble of primate hippocampal cells. *Journal of Neurophysiology*, 79:1797–1813, 1997b.

E. T. Rolls, J. Xiang, and L. Franco. Object, space, and object-space representations in the primate hippocampus. *Journal of Neurophysiology*, 94: 833–844, 2005.

E. T. Rolls, S. M. Stringer, and T. Elliot. Entorhinal cortex grid cells can map hippocampal place cells by competitive learning. *Network: Computation in Neural Systems*, 447:447–465, 2006.

A. Rotenberg and R. U. Muller. Variable place-cell coupling to a continuously viewed stimulus: evidence that the hippocampus acts as a perceptual system. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 352:1505–1513, 1997.

D. L. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Physical Review Letters*, 73:814–817, 1994.

F. Sargolini, M. Fyhn, T. Hafting, B. L. McNaughton, M. P. Witter, M.-B. Moser, and E. I. Moser. Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science*, 312(5774):758–762, 2006.

E. Save, A. Cressant, C. Thinus-Blanc, and B. Poucet. Spatial firing patterns of hippocampal place cells in bind rats. *Journal of Neuroscience*, 18:1818–1826, 1998.

B. Schoelkopf, S. Mika, C. J. C. Burgess, P. Knirsch, K. Muller, G. Rätsch, and A. J. Smola. Input space versus feature space in kernel-based methods. *IEEE Transactions on Neural Networks*, 10:1000–1017, 1999.

E. P. Sharp. Computer simulation of hippocampal place cells. *Psychobiology*, 19(2):103–115, 1991.

E. P. Sharp, H. T. Blair, and J. Cho. The anatomical and computational basis of the rat head-direction cell signal. *Trends in Neurosciences*, 24(5): 289–294, 2001.

P. E. Sharp. Multiple spatial/behavioral correlates for cells in the rat postsubiculum: Multiple regression analysis and comparison to other hippocampal areas. *Cerebral Cortex*, 6:238–259, 1996.

P. E. Sharp, J. Kubie, and R. Muller. Firing properties of hippocampal neurons in a visually symmetrical environment: Contributions of multiple sensory cues and mnemonic properties. *Journal of Neuroscience*, 10:3093–3105, 1990.

W. E. Skaggs and B. L. McNaughton. Spatial firing properties of hippocampal CA1 populations in an environment containing two visually identical regions. *Journal of Neuroscience*, 18(20):8455–8466, 1998.

W. E. Skaggs, J. J. Knierim, H. S. Kudrimoti, and B. L. McNaughton. A model of the neural basis of the rat's sense of direction. *Advances in neural information processing systems (NIPS)*, 7:173–180, 1995. doi: doi:10.1371/journal.pcbi.0030112.

T. Solstad, E. I. Moser, and G. T. Einevoll. From grid cells to place cells: A mathematical model. *Hippocampus*, 2006.

E. Y. Song, Y. B. Kim, Y. H. Kim, and M. W. Jung. Role of active movement in place-specific firing of hippocampal neurons. *Hippocampus*, 15:8–17, 2005.

H. Sprekeler, C. Michaelis, and L. Wiskott. Slowness: An objective for spike-timing-plasticity? *PLoS Computational Biology*, 3(6):e112, 2007.

R. W. Stackman and J. S. Taube. Firing properties of rat lateral mammillary single units: head direction, head pitch and angular head velocity. *Journal of Neuroscience*, 18(21):9020–9037, 1998.

R. W. Stackman and M. B. Zugaro. Self-motion cues and resolving intermodality conflicts: Head direction cells, place cells, and behavior. In S. I. Wiener and J. S. Taube, editors, *Head direction cells and the neural mechanisms of spatial orientation*, chapter 7, pages 137–162. MIT Press, 2005.

R. W. Stackman, E. J. Golob, J. P. Bassett, and J. S. Taube. Passive transport disrupts directional path integration by rat head direction cells. *Journal of Neurophysiology*, 90:2862–2874, 2003.

H.-A. Steffenach, M. Witter, M.-B. Moser, and E. Moser. Spatial memory in the rat requires the dorsolateral band of the entorhinal cortex. *Neuron*, 45(2):301–313, 2005.

J. V. Stone and A. Bray. A learning rule for extracting spatio-temporal invariances. *Network: Computation in Neural Systems*, 6:429–436, 1995.

P. Stopka and D. W. Macdonald. Way-marking behaviour: an aid to spatial navigation in the wood mouse (apodemus sylvaticus). *BMC Ecology*, 3:3, 2003.

S. M. Stringer and E. T. Rolls. Invariant object recognition in the visual system with novel views of 3D objects. *Neural Computation*, 14:2585–2596, 2002.

S. M. Stringer, G. Perry, E. T. Rolls, and J. H. Proske. Learning invariant object recognition in the visual system with continuous transformations. *Biological Cybernetics*, 94:128–142, 2006.

R. Tamura, T. Ono, M. Fukuda, and K. Nakamura. Spatial responsiveness of monkey hippocampal neurons to various visual and auditory stimuli. *Hippocampus*, 2(3):307–322, 1992.

H. Tanila. Hippocampal place cells can develop distinct representations of two visually identical environments. *Hippocampus*, 9:235–246, 1999.

H. Tanila, P. Sipila, M. Shapiro, and H. Eichenbaum. Brain aging: Impaired coding of novel environmental cues. *Journal of Neuroscience*, 17:5167–5174, 1997.

J. S. Taube and J. P. Bassett. Persistent neural activity in head direction cells. *Cerebral Cortex*, 13:1162–1172, 2003.

J. S. Taube, R. U. Muller, and J. B. Ranck Jr. Head direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *Journal of Neuroscience*, 2(10):420–435, 1990.

L. T. Thompson and P. J. Best. Place cells and silent cells in the hippocampus of freely-behaving rats. *Journal of Neuroscience*, 9(7):2382–2390, 1989.

S. Thorpe. Localized versus distributed representations. In *The Handbook of Brain Theory and Neural Networks*, pages 643–646. MIT Press, 2003.

S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381(6582):520–522, 1996.

E. C. Tolman. *Purposive Behavior in Animals and Men*. The Century Co., New York, 1932.

E. C. Tolman. Cognitive maps in rats and man. *Psychological Review*, 55: 189–208, 1948.

Toucan Corporation. Toucan virtual museum. http://toucan.web.infoseek.co.jp/3DCG/3ds/FishModelsE.html, 2005.

D. S. Touretzky. Attractor network models of head direction cells. In S. I. Wiener and S. Taube, editors, *Head direction cells and the neural mechanisms of spatial orientation*, chapter 18, pages 411–432. MIT Press, 2005.

R. Turner and M. Sahani. A maximum-likelihood interpretation for slow feature analysis. *Neural Computation*, 19(4):1022–1038, 2007.

N. Uanovsky and C. F. Moss. Hippocampal cellular and network activity in freely moving echolocating bats. *Nature Neuroscience*, 10(2), 2007.

S. Ullman and E. Bart. Recognition invariance obtained by extended and invariant features. *Neural Networks*, 17:833–848, 2004.

J. H. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society London B*, 265:359–366, 1998.

R. Vogels. Effect of image scrambling on inferior temporal cortical responses. *NeuroReport*, 10:1811–1816, 1999.

G. Wallis and E. T. Rolls. Invariant face and object recognition in the visual system. *Progress in Neurobiology*, 51(2):167–194, 1997.

S. I. Wiener and S. Taube, editors. *Head direction cells and the neural mechanisms of spatial orientation*. MIT Press, 2005.

S. I. Wiener, C. A. Paul, and H. Eichenbaum. Spatial and behavioral correlates of hippocampal neuronal activity. *Journal of Neuroscience*, 9(8): 2736–2763, 1989.

Wikipedia. Grid cells — wikipedia, the free encyclopedia, 2007. URL `http://en.wikipedia.org/w/index.php?title=Grid_cells&oldid=165890133`. [Online; accessed 22-November-2007].

B. Willmore and D. J. Tolhurst. Characterizing the sparseness of neural codes. *Network: Computation in neural systems*, 12(3):255–260, 2001.

T. J. Wills, C. Lever, F. Cacucci, N. Burgess, and J. O'Keefe. Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308:873–876, 2005.

D. J. Willshaw, O. P. Buneman, and H. C. Longuet-Higgins. Non-holographic associative memory. *Nature*, 222:960–962, 1969.

M. A. Wilson and B. A. McNaughton. Dynamics of the hippocampal ensemble code for space. *Science*, 261(5124):1055–1058, 1993. Erratum in Science 1994 Apr 1;264(5155):16.

L. Wiskott. Learning invariance manifolds. In L. Niklasson, M. Bodén, and T. Ziemke, editors, *Proceedings of the 8th International Conference on Artificial Neural Networks, ICANN'98, Skövde*, Perspectives in Neural Computing, pages 555–560, London, 1998. Springer.

L. Wiskott. Slow feature analysis: A theoretical analysis of optimal free responses. *Neural Computation*, 15(9):2147–2177, 2003.

L. Wiskott. How does the visual system achieve shift and size invariance? In J. L. van Hemmen and T. J. Sejnowski, editors, *23 Problems in systems neuroscience*. Oxford University Press, 2004.

L. Wiskott and T. Sejnowski. Slow feature analysis: Unsupervised learning of invariances. *Neural Computation*, 14(4):715–770, 2002.

D. P. Wolfer and H. P. Lipp. Dissecting the behaviour of transgenic mice: is it the mutation, the genetic background, or the environment? *Experimental Physiology*, 85(6):627–634, 2000.

E. R. Wood, P. A. Dudchenko, and H. Eichenbaum. The global record of memory in hippocampal neuronal activity. *Nature*, 397(6720):613–6, 1999.

E. R. Wood, P. A. Dudchenko, R. J. Robitsek, and H. Eichenbaum. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*, 27:623–633, 2000.

R. Wyss, P. König, and P. Verschure. A model of the ventral visual system based on temporal stability and local memory. *PLoS Biology*, 4(5):e120, 2006.

D. Yoganarasimha, X. Yu, and J. J. Knierim. Head direction cell representations maintain internal coherence during conflicting proximal and distal cue rotations: Comparison with hippocampal place cells. *Journal of Neuroscience*, 26(2):622–631, 2006.

B. J. Young, G. D. Fox, and H. Eichenbaum. Correlates of hippocampal complex-spike cell activity in rats performing a nonspatial radial maze task. *Journal of Neuroscience*, 14:6553–6563, 1994.

M. B. Zugaro, A. Arleo, A. Berthoz, and S. I. Wiener. Rapid spatial reorientation and head direction cells. *Journal of Neuroscience*, 23(8):3478–3482, 2003.

M. B. Zugaro, A. Arleo, C. Déjean, E. Burguière, M. Khamassi, and S. I. Wiener. Rat anterodorsal thalamic head direction neurons depend upon dynamic visual signals to select anchoring landmark cues. *European Journal of Neuroscience*, 20:530–536, 2004.

Abbreviations

Abbreviation	Meaning
CA	Cornu Ammonis
CL	Competitive Learning
DG	Dentate Gyrus
EC	Entorhinal Cortex
ICA	Independent Component Analysis
IT	Inferotemporal Cortex
PCA	Principal Component Analysis
SFA	Slow Feature Analysis
SFP	Spatial Firing Pattern
V1	Primary Visual Cortex

Acknowledgements

Scientists consider themselves as dwarfs standing on the shoulders of giants and hope to look a bit farther than their predecessors. Luckily I didn't stand there alone during the work on my thesis but found a merry crowd around me at the ITB.
The ITB feels very much like home after these four years and I want to thank all the people who are or were part of it.
Especially, I want to thank my supervisor Laurenz Wiskott who always had time to discuss ideas and problems, for highly concentrated discussions, his great intuition, fair treatment and for putting only minimal teaching and administrative burdens on me.
Without my colleagues and friends at the institute I might not have finished this thesis – thanks to Andreas, Jan, Henning, Martin, Niko, Pietro, Raphael, Roland, Samuel, Susanne, Tiziano, Tobias, and everybody else at the institute!
Special thanks go to Richard Kempter and Christian Leibold for many answers about the hippocampus. Extra thanks go to Pietro and Tiziano for convincing me of Python as a programming language, their great MDP toolbox, and for constant help in related questions. Big thanks to Henning for theoretical support whenever I felt the need. He comes close to our proverbial Russian mathematician who already solved all our complex problems as a homework exercise years before. Thanks also to the Volkswagen Foundation for funding the junior research group that I was part of.

Lebenslauf

Mathias Franzius	
6.3.1975	Geboren in Minden
1995 – 2003	Studium der Informatik an der Brandenburgischen Technischen Universität Cottbus
2003	Diplomarbeit am Institut für Neuroinformatik an der ETH Zürich bei Prof. König
5.2003 – 10.2007	Wissenschaftlicher Mitarbeiter (später Gastwissenschaftler) an der Humboldt-Universität zu Berlin bei Prof. Wiskott, Institut für Biologie

Veröffentlichungen

Publikationen

M. Franzius, R. Vollgraf, L. Wiskott: From Grids to Places, J Comput Neurosci (2007) 22:297–299

M. Franzius, H. Sprekeler, L. Wiskott: Slowness and Sparseness Lead to Place, Head-Direction, and Spatial-View Cells. PLoS Comput Biol (2007) 3(8): e166. doi:10.1371/journal.pcbi.0030166

Konferenzbeiträge

M. Franzius, H. Sprekeler, L. Wiskott: Unsupervised learning of place cells and head direction cells with slow feature analysis. Proc. 7th Meeting of the German Neuroscience Society - 31st Göttingen Neurobiology Conference, Göttingen (2007)

M. Franzius, H. Sprekeler, L. Wiskott: Unsupervised learning of visually driven place cells in the hippocampus, Beiträge zur 8. Jahrestagung der Gesellschaft für Kognitionswissenschaft, Saarbrücken, eds. C. Frings, A. Mecklinger, B. Opitz, M. Pospeschill, D. Wentura, H.D. Zimmer, publ. Shaker Verlag, Aachen, p. 60 (2007)

M. Franzius, H. Sprekeler, L. Wiskott: Slowness leads to place cells, Proc. Berlin Neuroscience Forum, Bad Liebenwalde, publ. Max-Delbrück-Centrum für Molekulare Medizin (MDC), Berlin, p. 42 (2006)

M. Franzius, H. Sprekeler, L. Wiskott: Slowness leads to Place Cells. Computational Neuroscience Conference (CNS), Edinburgh (2006)

W. Einhäuser, C. Kayser, M. Franzius, J. Hipp, G.U. Moeller K.P. Körding and P. König: Learning from Natural Videos: From Complex Cells to object classification. Annual Meeting of the Swiss Society for Neuroscience, Lausanne (2004)

W. Einhäuser, C. Kayser, M. Franzius, J. Hipp, K.P. Körding and P. König: Learning from the temporal statistics of natural scenes: From V1 properties to object classification. Meeting of the Center for Neuroscience Zürich (2003)

M. Franzius, K.P. Körding, P. König: A hierarchical model of cortical function learns texture recognition. Meeting of the Center for Neuroscience Zürich (2002)

Erklärungen

Hiermit erkläre ich, die vorliegende Arbeit selbständig ohne fremde Hilfe verfaßt und nur die angegebene Literatur und Hilfsmittel verwendet zu haben. Die Promotionsordnung ist mir bekannt. Hiermit erkläre ich weiterhin, dass ich bisher keinen Doktorgrad besitze und mich nicht anderwärts um einen Doktorgrad beworben habe.

Mathias Franzius
15.11.2007